

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 2, 2004, 15:00:02 ; Search time 13693 Seconds  
(without alignments)  
11357.249 Million cell updates/sec

Title: US-10-780-347-1

Perfect score: 3588

Sequence: 1 gtgcattattgcatgatg.....gatnaaattgagcggcttaa 3588

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*\*

1: gb\_ba:\*\*

2: gb\_htg:\*\*

3: gb\_in:\*\*

4: gb\_om:\*\*

5: gb\_ov:\*\*

6: gb\_pat:\*\*

7: gb\_ph:\*\*

8: gb\_pl:\*\*

9: gb\_pr:\*\*

10: gb\_ro:\*\*

11: gb\_sts:\*\*

12: gb\_sy:\*\*

13: gb\_un:\*\*

14: gb\_vi:\*\*

15: gb\_ba:\*\*

16: em\_fun:\*\*

17: em\_hum:\*\*

18: em\_in:\*\*

19: em\_mu:\*\*

20: em\_om:\*\*

21: em\_or:\*\*

22: em\_ov:\*\*

23: em\_pat:\*\*

24: em\_ph:\*\*

25: em\_pl:\*\*

26: em\_ro:\*\*

27: em\_sts:\*\*

28: em\_un:\*\*

29: em\_vi:\*\*

30: em\_htg\_hum:\*\*

31: em\_htg\_inv:\*\*

32: em\_htg\_other:\*\*

33: em\_htg\_mus:\*\*

34: em\_htg\_pln:\*\*

35: em\_htg\_rod:\*\*

36: em\_htg\_man:\*\*

37: em\_htg\_vrt:\*\*

38: em\_sy:\*\*

39: em\_htgo\_hum:\*\*

40: em\_htgo\_mus:\*\*

41: em\_htgo\_other:\*\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	108.8	3.0	302100	1	AP005087	AP005087 Vibrio pa
C 2	105	2.9	248850	1	AP005348	AP005348 Vibrio vu
C 3	105	2.9	300883	1	AE016809	AE016809 Vibrio vu
C 4	99.4	2.8	300395	1	AE016785	AE016785 Pseudomon
C 5	91.6	2.6	301457	1	AE016924	AE016924 Chromobac
C 6	82.8	2.3	190050	1	AL646080	AL646080 Ralstonia
C 7	80	2.2	31154	1	AF361470	AF361470 Rhizobium
C 8	79.4	2.2	14867	3	AE001398	AE001398 Plasmidiu
C 9	76.8	2.1	252394	3	AE014833	AE014833 Plasmidiu
C 10	76.2	2.1	179902	2	AP003630	AP003630 Oryza sat
C 11	75.4	2.1	143409	9	AL162497	AL162497 Human DNA
C 12	74.8	2.1	164188	2	EX005205	EX005205 Danio rer
C 13	74.6	2.1	110000	2	PFMAL7P1_07	Continuation (8 of
C 14	74.6	2.1	170627	2	AC125567	AC125567 Rattus no
C 15	74.6	2.1	182518	9	AC131649	AC131649 Homo sapi
C 16	73.2	2.0	168885	9	AC098691	AC098691 Homo sapi
C 17	73	2.0	83703	5	AL732635	AL732635 Zebrafish
C 18	72.8	2.0	11422	6	AX345121	AX345121 Sequence
C 19	72.8	2.0	11422	6	AX348323	AX348323 Sequence
C 20	72.8	2.0	223021	2	AC119799	AC119799 Pan trogl
C 21	72.8	2.0	34980	6	AX344564	AX344564 Sequence
C 22	72.6	2.0	5727	1	AF037441	AF037441 Edwardsie
C 23	72.4	2.0	8093	6	AX598978	AX598978 Sequence
C 24	72.4	2.0	8093	6	AX767538	AX767538 Sequence
C 25	72.4	2.0	8093	6	AX795912	AX795912 Sequence
C 26	72.4	2.0	8093	6	AX822432	AX822432 Sequence
C 27	72.4	2.0	8093	6	AX826072	AX826072 Sequence
C 28	71.8	2.0	165228	5	AL929341	AL929341 Zebrafish
C 29	71.4	2.0	61020	6	AX251546	AX251546 Sequence
C 30	71.4	2.0	131033	5	AL591370	AL591370 Zebrafish
C 31	71.2	2.0	150865	5	AL929188	AL929188 Zebrafish
C 32	71.2	2.0	208140	2	EX255953	EX255953 Danio rer
C 33	70.8	2.0	13568	1	AE009363	AE009363 Agrobacte
C 34	70.8	2.0	19961	1	AE008251	AE008251 Agrobacte
C 35	70.8	2.0	110000	2	PFMAL6P1_02	Continuation (3 of
C 36	70.8	2.0	175026	5	AL954861	AL954861 Zebrafish
C 37	70.8	2.0	179435	2	AC134381	AC134381 Danio rer
C 38	70.8	2.0	229480	5	AL929150	AL929150 Zebrafish
C 39	70.6	2.0	10029	1	AE013970	AE013970 Yersinia
C 40	70.6	2.0	73243	9	AL731858	AL731858 Human DNA
C 41	70.6	2.0	194869	5	AL954136	AL954136 Zebrafish
C 42	70.6	2.0	201050	1	AJ414143	AJ414143 Yersinia
C 43	70.2	2.0	101034	9	AC098550	AC098550 Homo sapi
C 44	70.2	2.0	258658	3	AE014832	AE014832 Plasmidiu
C 45	69.8	1.9	8666	6	AX767525	AX767525 Sequence

# ALIGNMENTS

RESULT 1  
AP005087/c  
LOCUS AP005087 302100 bp DNA linear BCT 05-MAR-2003  
DEFINITION Vibrio parahaemolyticus DNA, chromosome 2, complete sequence, 4/6.  
ACCESSION AP005087 BA000032  
VERSION AP005087.1 GI:28809161  
KEYWORDS  
SOURCE Vibrio parahaemolyticus  
ORGANISM Vibrio parahaemolyticus  
Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
Vibrionaceae; Vibrio.  
REFERENCE  
AUTHORS Nasu, H., Iida, T., Sugahara, T., Yamaichi, Y., Park, K.S., Yokoyama, K.,  
Makino, K., Shinagawa, H. and Honda, T.  
TITLE A filamentous phage associated with recent pandemic Vibrio

Tue Jul 6 09:50:11 2004

```

paraHaemolyticus O3:K6 strains
J. Clin. Microbiol. 38 (6), 2156-2161 (2000)
20295086
10834969
REFERENCE
2
AUTHORS
Makino, K., Oshima, K., Kurokawa, K., Yokoyama, K., Uda, T.,
Tagomori, K., Iijima, Y., Najima, M., Nakano, M., Yamashita, A.,
Kubota, Y., Kimura, S., Yasunaga, T., Honda, T., Shinagawa, H.,
Hattori, M. and Iida, T.
Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
distinct from that of V. cholerae
Lancet 361 (9359), 743-749 (2003)
JOURNAL MEDLINE
22508454
PUBMED
12620739
3 (bases 1 to 302100)
Oshima, K., Kurokawa, K., Makino, K., Yokoyama, K., Yasunaga, T.,
Honda, T., Shinagawa, H., Hattori, M. and Iida, T.
Direct Submission
Submitted (09-APR-2002) Ken Kurokawa, Osaka University, Genome
Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail: ken@gen-info.osaka-u.ac.jp,
URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365,
Fax: 81-6-6879-2047)
genome project
This clone was isolated from a patient presenting with acute
gastroenteritis.
FEATURES
source
Location/Qualifiers
1. 302100
/organism="Vibrio parahaemolyticus"
/mol_type="genomic DNA"
/strain="O3:k6"
/sub_strain="BIMD 2210633"
/db_xref="taxon:670"
/chromosome="2"
53..1139
misc_feature
/note="disrupted by frameshift, similar to part of
hypothetical protein GB:AAF96452.1 (AE004386) percent
identity 62 in 165 aa"
complement (1224..2012)
/gene="VPA0861"
complement (1224..2012)
/gene="VPA0861"
/note="similar to GB:AAF96483.1 (AE004389) percent
identity 73 in 261 aa"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAC62204.1"
/db_xref="GI:28809162"
/translation="MKVWIKASALLTLAGLTACASPTGRNQILMFSDQMSSLGAKS
PDOMKDIPIISKDKTNAVQCVAKQITNAIPPOAGFKDWVVFVDSQVNAFALPGG
KIGVYTGLLKVAQNQDLATVIGHEVAHLAEHSNERLSOLANAGLSLANVAGAS
EYKQYQLTWAALGVQYGVILPYGPTQSEADIVGLEVMKAGFNQSVDLWKNM
SAASGAGQPPFFSTHPSHSTRIKDLQATINKPKYNVAKPKCG"
2217..2651
/gene="VPA0862"
2217..2651
/gene="VPA0862"
/note="similar to GB:AAF96482.1 (AE004389) percent
identity 75 in 144 aa"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAC62205.1"
/db_xref="GI:28809163"
/translation="MSLWKPKIPISLDILNATSRNTMWEHLOIEVNTFTDDSLSATMPVC
SFTQPIGLMGGASVLAELTGLSIAANFCVSEGNVCGLDINAHVRAKESGYVIGT
AKPIHLGVSTQWQVLEITDERDRUVCTSRUTIAVRQQAALA"
2694..2927
/gene="VPA0863"
2694..2927
/gene="VPA0863"

```

```

/gene="VPA0863"
/note="similar to GB:AAF96481.1 (AE004389) percent
identity 56 in 76 aa"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC62206.1"
/db_xref="GI:28809164"
/translation="MWIEFSLGKIIATQREIVIKLTGEMVMTLQAQTDALQLLORGAN
VLAHGAECRWSIKIDNEEQLIQAQBIGIDIQ"
3007..4713
/gene="VPA0864"
3007..4713
/gene="VPA0864"
/note="similar to GB:AAF96480.1 (AE004388) percent
identity 66 in 565 aa"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative ABC transporter periplasmic solute
binding protein"
/protein_id="BAC62207.1"
/db_xref="GI:28809165"
/translation="MSSPRLRVQFETLFEKAGNTEVQLEDITDALFCTRRNARIVL
NKLBEEGHIEWHPAAGRGKSLIFKFNRSVSENLAARYLDEKIQALDLDNDAA
RLTQVIOGLQHRQGEVVRPLPYRPLAMLNPKPMRSEQHIARQVFSGLTRDE
SEQLQDLAHTWEALSDTHWFLRGRVRHNGBELTDCVVEISILALGNLFSHIQ
QVSSQPWTVDIKLVPRDRIPLALSSQAKVLLQALRSEDFRKPITGPGFOVKYN
DDKRLITAFQYFGFRLGLQGEVWVIDEAYSMVPSLSKPKMKQAQSTDEVELDP
GCTFLLNKNTGIAGKPRWAEFLSQTLNSHOIYAHVQDQKVIELGVLOAFELKPGWID
LRPT"SVGEVPOQKSIITVAYQKHPMPFVIAKAIKHLKPHSDVEFIRYDTQPNPE
EVDVWVKANGIATNRDALGWLIDYDIKFSAGYDFSEWATLDQWRAGIHADPFA
RELGRQLVKSCQVIMPFHCLWGLGNKDHSGALQNAKNAKLGWDFSNVWVKPEIDNGQ
AE"
complement (5251..5847)
/gene="VPA0865"
complement (5251..5847)
/gene="VPA0865"
/note="similar to GB:AAF96687.1 (AE004407) percent
identity 71 in 192 aa"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAC62208.1"
/db_xref="GI:28809166"
/translation="MKSTSTLIGLTIYVAIVFLFSALEPPSSRAVWFABEIIIPAIGILIA
TWAISIRYQNSNTAYLLMFIWLCIHTIGAKVTFAEVPDFNPNLIGSRNNFDRVAHF
AIGLYAYFIASLYLRNKXFNPTFCWFALFAIMSLAAGYEIIWWYAEALAGDBGIAF
LGSQDIDWAQDKMLCDITIGAILSLFLMSAORRPAKPF"
6123..6962
/gene="VPA0866"
6123..6962
/gene="VPA0866"
/note="similar to GB:AAK22187.1 (AE005694) percent
identity 30 in 298 aa"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative oxidoreductase iron/ascorbate family"
/protein_id="BAC62209.1"
/db_xref="GI:28809167"
/translation="MKLETVDYLADDAQQQVASIRETGFGVLKNHPIPOELVESIYK
NMHEFFCSEKNDQFQNVETQYGFPPSPSVAGKHVKDKIKEYHYTPMGQIPBOLK
ABILEYPRANAPAEQLIGWVEYAPKQVSKFSALEMINNSKTLRLVLRHPMT
GESEFGAIAAAHEDINLLTPAANFGLQWNSGEMIDVPCDFGNLIINIGMDLQ
BASGGYFSTTHRVINPTGARQEKSRISLPLFLHPKPTVLSERYTAHSYLMERLEL
GVI"
7095..7715
/gene="VPA0867"
7095..7715
/gene="VPA0867"

```

```

/note="similar to GB:AAF96686.1 (AE004407) percent
identity 63 in 204 aa"
/codon_start=1
/evidence=not experimental
/transl_table=1
/product="DnaJ-related protein"
/protein_id="BAC62210.1"
/db_xref="GI:28809168"
/translation="MEQYGVCSQSEIENPLLWMLFIILKRPQSGWKVHTLASKLS
ERGYINLQDPEDKDLKPRNFLLMNALYQIETLFPDSMLQVAMDIMLSALEASRN
SYVDNPEDQLQWSYEDNEVRLLNEFWTRQFVGGAASEHNDKAKALKLFD
LAGDYSRTEIRQWRKLALRHPDRENGDAERFVLCEAWNVLNRG"
/complement(7844..9499)
/gene="VPA0868"
/complement(7844..9499)
/gene="VPA0868"
/complement(7844..9499)
/note="similar to GB:AAF96686.1 (AE004407) percent
identity 78 in 548 aa"
/codon_start=1
/evidence=not experimental
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="BAC62211.1"
/db_xref="GI:28809169"
/translation="MDQTATLTKIEKONVRAVQKIQGYDFTDFTLIDHVQGDPIYA
SASRFRATRWSLTGLKLEWDESPAFORARDFIARSPQAKENTVSIANGOTVL
DSYLVLTTEGIEIRFVRNPAEGRSVLGKKANNILFHLKPIRATRELDSQVM
VKHQVVEDQALREQLAEHLNVAFGVANGDLPMKCAVEFTAPESLQVT
LHAPKGVVTGLGT;KGTLLTGMGDFHKTLLNATERSIYDHI PDQGREY;VTQDKA
MKI RAEKRGCRVHLNLSYINHL PMGKDTADFSTODASGTSQAQWLQSI EAGATSL
LIDETSATNEMI DERMOALVAGDEPITPLVDRIQLDRDELDTSLIIVMGSGDYL
DVANVTIOHDIQAVDVEKAKVIOAHPTQRNNESES;LQTFPRALNRVALMMLLT
DKRVKSGADYRLEKREFIDLSALQSI ESDAVNAIGLWFLQAQLPGWCNNPAKE
IEMUSGWSHASLPKQGLAKPRFTLDYNAALNMRKSPKPSH"
/complement(9571..11481)
/gene="VPA0869"
/complement(9571..11481)
/gene="VPA0869"
/note="similar to GB:AAF96683.1 (AE004406) percent
identity 50 in 612 aa"
/codon_start=1

Query Match          3.0%; Score 108.8; DB 1; Length 302100;
Best Local Similarity 59.6%; Pred. No. 1.9e-12;
Matches 201; Conservative 0; Mismatches 133; Indels 3; Gaps 1;

QY 3246 ACTCACAACATGAAGCGCTTAGTAGAACCGCTGCAGCATCCTATGACGTGAACGA 3305
Db 192447 ACTCGAACTTTCACGAGTACGGAAGCCTCGTGTCACATCTACTAGTGTGGAACCG 192388
QY 3306 ATGGCGCCGCTAAAGACGAAGAGCTGCCGTTTGTGTGTGCGCTCATTGCGCACTTTCAG 3365
Db 192387 AAGGAACAACCTTTGAAAAAAGAGCTTCCTTTTGTGTGGGTGTATTGCGGCACTTTCGAG 192328
QY 3366 GACACAAACAGCAATCAGAAAAGCTTGATTATAGAGACGCGAGAGTTCACGGGTATCGATA 3425
Db 192327 GCCAAATACCGAAGCATTTGAAA----CAATTAAGGATCGCGGTTTCATTCAAATCGATC 192271
QY 3426 AAGACAACCTTCGATACAGTGATGGGGCAAAATTCACCGCGCTTTTCGTACAAGGTTGATA 3485
Db 192270 GTGATAACTTCGATGATGTACTTAAAAAGATGAGTCATCGGTTAAGTTCAAAGTCGCAA 192211
QY 3486 ACAAGCTTCCTAATGATGATGCCAGTTTGAAGTGAAGTGACCTGAGCCTCCGTTCCATCAAG 3545
Db 192210 ACAAGTTAGCAATGACGATTCCTGAATTTTCGGGTTGAAC TGAGCTTTTAAATCGATCAAG 192151
QY 3546 ATTTCACCCACGAGAACCTTAGTTGATNAAATTGAGCC 3582
Db 192150 ACTTTCAACACGACGAGATTGTTCATCAAGTAGAAC 192114

```

```
QDPKLIKLOAQYWOAAKEWQOALELWRIAKQBOSQYQWEVALLLENOLQRYQVQLISLA
EMDNPKQFPAQLLKINALYRLERIDQAYAAKAKADALRSTKQTQNWLFSLSHKLKSS
ESKLDKAGQGVSEREHLTQORNVADKSDSSS"
/complement(3182. .3820)
/gene="VVA0916"
/complement(3182. .3820)
/gene="VVA0916"
/codon_start=1
/transl_table=11
/product="periplasmic protein TonB2"
/protein_id="BAC96942.1"
/db_xref="GI:37201119"
/translation="MMKLAFVVALALVFALFSAALLVNSKKQAIQDTSPSLSNM
VWVQDTEVTRRSQSPPEIPEPPQPTKMTQPMQVTSQVATPKIDMRFNVSQ
LAVSIPTAVSAPTQASNTTKQLMPTSRVEAVYPAKAKKRIEYVILKFDIDTQRT
KNIEVPAQPARPFERNAMDAIKWRYCPQVDGQAQTIQGYTKIEFKMAQ"
/complement(3820. .4224)
/gene="VVA0917"
/complement(3820. .4224)
/gene="VVA0917"
/codon_start=1
/transl_table=11
/product="biopolymer transport protein ExbD2"
/protein_id="BAC96943.1"
/db_xref="GI:37201120"
/translation="MRUGRRQSKSDEAIDLTSMLDIVFIMLFFIVTSSFVRESQVE
VNRQAAVSAQKADGFVATTAANDIYIDKRMVDVERVQASIERLLLEQPDASLVIQ
ADEYANGVVVYKMDAKAAGVKNIAAEKS"
/complement(4221. .4787)
/gene="VVA0918"
/complement(4221. .4787)
/gene="VVA0918"
/codon_start=1
/transl_table=11
/product="biopolymer transport protein ExbE2"
/protein_id="BAC96944.1"
/db_xref="GI:37201121"
/translation="MIIMDIHSDIWPPELLTSSMMNALYSFMQGGVGLWMLAAVAVI
FWLITIKLFIYCAYRLRAETLQKQMRHDLHSFARSQREAWLDDAQQLFKHLA
FIVLVAICPKLGLLGTVMGISVFDVMAQTGSSNPKLMASGISLATLPTMGWMAAL
VGMFVHARLVKMLQREVLQSLRSDS"
/complement(4784. .6151)
/gene="VVA0919"
/complement(4784. .6151)
/gene="VVA0919"
/codon_start=1
/transl_table=11
/product="biopolymer transport protein ExbB-related
protein"
/protein_id="BAC96945.1"
/db_xref="GI:37201122"
/translation="MNKRLISLALIVLSLPSFAOTQLVPKATQSRQQAENQORE
NAFLSQOQLAEKRLQEKQKSTLEKNALLSQFSDNEVTLSHLEELRVEVTSIGE
LFGVGRQALKEADLNDVATIGTEQVETVQIVAAQSPDPIAQKILHWVHQQI
LVGMERQALQVNIAGAEVPTDVNAGNFALLAQGFVNWKSXGFAIPAKDEG
APTLDTRSHATSMVPLIDPTRGVLLQEQYHPSLADRFAGGVGVGVIFALLAVG
LVIALYRGILLITTKIRLQKPNQGNPLGRVLAVVDKPKRSPVLEIRLLET
IVDQTCLEGLSKMLKLAAPMLGLLGTVMGIVTFOVITQFNGDKPKVWASGISM
ALVTVLGLVNAEMLLLAHNVLSGWAABEIRLSILEKQIGLVAQASQQLPSLTPTMDK
VGNAA"
/complement(6152. .6919)
/gene="VVA0920"
/complement(6152. .6919)
/gene="VVA0920"
/translation="MSGYKAGALLWVAFGMSNDLEQAAQIKQNRHTAASKSQSV
INASDKSQLQAEIILKEEVNLELYQSHLTSMTANQOEKSLAKQIQIVTRQ
GIVPLMYMDLDTLKTINQDMPIKWQQRQBERIEKLEQVMVADVADAEKFRILEAYQ
/complement(identified by GeneMark and Glimmer2"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAC96946.1"
/db_xref="GI:37201123"
/translation="MSGYKAGALLWVAFGMSNDLEQAAQIKQNRHTAASKSQSV
INASDKSQLQAEIILKEEVNLELYQSHLTSMTANQOEKSLAKQIQIVTRQ
GIVPLMYMDLDTLKTINQDMPIKWQQRQBERIEKLEQVMVADVADAEKFRILEAYQ
```

```
IELDYGSKLSAYQAPITLADGKQIEVEQLHLGRVSLIARTLNANQFWNDONQRHWS
LTTIPSTDVNAFDIAAKQVAPMLNMFVALATAEK"
/complement(7182. .7670)
/gene="VVA0921"
/complement(7182. .7670)
/gene="VVA0921"
/codon_start=1
/transl_table=11
/product="uncharacterized conserved protein"
/protein_id="BAC96947.1"
/db_xref="GI:37201124"
/translation="MRVWTSKVAFVDVQNIYTCCKEYRKHFDYNEFWSQATQDKOV
VWRAIYAISSKDPQCFQFHLIRGIGEDVKLKPFIQWDSAGKDWGLADAFEAH
PMWDEVILLSDGDFEILVERIQQRKVKVTVIGVPLTAQNLOAVVDKPIPIEDEL
LP"
/7796. .8227
/gene="VVA0922"
/7796. .8227
/gene="VVA0922"
/translation="identified by GeneMark and Glimmer2"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC96948.1"
/db_xref="GI:37201125"
/translation="MRVRSIVLILFSAHSYPADSEKMOHRSIMFAPSHDQVVOQF
LEIEMECSCQERDITVLVIKEDGSVPVWQADAPLSVLNAYFIDIPKQHTGILVG
KDGEKRLWGEKTDWTLKKVIDQMLPRQEMQNRPSRCAL"
/complement(8234. .9076)
/gene="VVA0924"
/complement(8234. .9076)
/gene="VVA0924"
/codon_start=1
/transl_table=11
/product="uroporphyrinogen-III methylase"
/protein_id="BAC96949.1"
/db_xref="GI:37201126"
/translation="MLIKSKHFPREALSGTLVHQIYISIGSTMTLTSLSSIQNGFVS
LVAGQGPDDLTVKGRVQOAEVVDRLVSPSEILALANAKAEIMYVGGKLDHFQV
QEBINQILVPALEKGRVVLKGDSPFIRGGGELEELAEHGVREVVPGITAAAG
ATATAGPLTHRHAQSVQFTIGHIQDKGEIEMPSLAQANNTLAVFMGLKQCRYIAE
KLENGLDASTSCAIENGTRRQORVLTGSLTQPELATQAIISPALIVVGSVTSLHK
LKWFL"
/8255. .8395
/gene="VVA0923"
/8255. .8395
/gene="VVA0923"
/translation="identified by Glimmer2"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC96950.1"
/db_xref="GI:37201127"
/translation="MQRSHADHDQRTDCLSGQFRELCOAARQYALLATGAIFFNRA
Query Match 2.9%; Score 105; DB 1; Length 248850;
Best Local Similarity 56.6%; Pred. No. 1.2e-11;
Matches 214; Conservative 0; Mismatches 161; Indels 3; Gaps 1;
QY 3209 ATGATATCTCATTTAGGACACCGTTATGGCATTGAACCTCACACATATAAGCGGTTAGT 3268
DB 95004 ATGATTAATTAAAGGAGAACGATATGACTAGTATCCATCGAATATTCGCGGATACGG 94945
QY 3269 AAGAACCGTGTACGATCACCTATGACGTTGAAACGAATGGCGCGCTTAAGACGAAAGAG 3328
DB 94944 AAGCCACGTTTCATATTACCTATGATGTTGAAAACAGAAGGCACATCACTCAGAGAAAG 94885
QY 3329 CTGCGGTTTGTGTGGCGTCATTGGCGACTTTTCAGCACACAAACAGAAATCAGAAAAA 3388
DB 94884 CTTCGGTTTGTAGTGGTGTGATGGGGATTTTCTGCGCCAAAATCTGAAGCCCTCAA 94825
QY 3389 GTTGATTTTAGAAGCGGAGAGTTTCACGGGTATCGATATAAGACAACTTCGATACAGTGATG 3448
```



```

Db 94824 ---CCACTAAAAGACCGCTGTTTATTCAAATTGATCGAGACAACTTGATGATGACTC 94768
QY 3449 GGGCAATTCACCGCGCTCTTTGTCACAAAGTTGATACAAAGCTTGCTATGATGATAGC 3508
Db 94767 AAGAAATAGTCCATCAGTGGAGTTTAAAGTGAAACACAGATGGTGGGATGACGGAACG 94708
QY 3509 CAGTTTGAAGTGAACCTTGAGCCCTCGGTTCCGATGAAGATTTCCACCCAGAGAACTTAGTT 3568
Db 94707 GAATTCGCGTCAATCTGCAGTTCCCAATCGATGAAGGATTTCCGACCCGCTCGATTGTT 94648
QY 3569 GATNAATTTGAGCGCTT 3586
Db 94647 AGACAAGTTGATCGGCTT 94630

RESULT 3
AE016809/c
LOCUS      AE016809          30883 bp      DNA      linear      BCT 29-SEP-2003
DEFINITION Vibrio vulnificus CMCP6 chromosome II section 2 of 6 of the
complete sequence.
ACCESSION  AE016809
VERSION    AE016796
KEYWORDS   AE016809.1  GI:27358288
SOURCE     Vibrio vulnificus CMCP6
ORGANISM   Vibrio vulnificus CMCP6
            Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
            Vibrionaceae; Vibrio.
REFERENCE  1 (bases 1 to 30883)
            Kim,Y.R., Lee,S.E., Kim,C.M., Kim,S.Y., Shin,E.K., Shin,D.H.,
            Chung,S.S., Choy,H.E., Progulste-Fox,A., Hillman,J.D., Handfield,M.
            and Rhee,J.H.
            Characterization and Pathogenic Significance of Vibrio vulnificus
            Antigens Preferentially Expressed in Septicemic Patients
            Infect. Immun. 71 (10), 5461-5471 (2003)
REFERENCE  2 (bases 1 to 30883)
            Jeong,H., Moon,Y.H. and Kim,J.J.
            Direct Submission
            Submitted (13-DEC-2002) Genotech Corp., 461-6, Jeonmin-dong,
            Yuseong-gu, Daejeon 305-811, South Korea
            3 (bases 1 to 30883)
            Rhee,J.H., Kim,S.Y., Chung,S.S., Lee,S.E. and Choy,H.E.
            Direct Submission
            Submitted (13-DEC-2002) Department of Microbiology, Genome Research
            Center for Enteropathogenic Bacteria, Chonnam National University
            Medical School, Hak-I-Dong, Dong-Gu, Kwang-Ju 501-746, South Korea
FEATURES             Location/Qualifiers
     source            1..300883
                        /organism="Vibrio vulnificus CMCP6"
                        /mol_type="genomic DNA"
                        /strain="CMCP6"
                        /db_xref="taxon:216895"
                        /chromosomes="II"
                        269..577
     gene              /locus_tag="VV20274"
                        269..577
     CDS               /locus_tag="VV20274"
                        /codon_start=1
                        /transl_table=11
                        /product="Unknown"
                        /protein_id="AAO07239.1"
                        /db_xref="GI:27358289"
                        /translation="MILENKPNTSRQFLISNNLKQCDFTYFHEKTRFOILKAQNTSTV
                        TISGKDFLKI GGVELTITQGRAIGICQRLAQEKDFVDASLRIADVKEACRVR
                        "
     gene              565..2553
                        /locus_tag="VV20275"
     CDS               565..2553
                        /locus_tag="VV20275"
                        /note="COG2199"
                        /codon_start=1
                        /transl_table=11
                        /product="FOG: GGDEF domain"

```

```

/protein_id="AAO07240.1"
/db_xref="GI:27358290"
/translation="MCSUKSLSKIKNKLILLPTGLTFLFLIAQGHFWASVEKQOOLQ
NHRISLAYGAQNLQASVFNQKVMQEIILDSFSDISDPSILILVELRSUEGGVLAQR
EMPVNDVHLSLPYFFETFTYQSRIQDEQNAIATSTATQTEQMQLRLESSV
VFLAILLSGLTYLIVQKVIDPVYKLNESMQDLVEGRIKGSLSVSHSSDEMGLVN
AFNAMSQALARDSDVHHTLRLEHKAFAFDVI EAVOTALLVNEHGDILHNAAR
MFGSDAORIAQHLSEMLCKEFSFVANALTSKNPLERLEVERCLPTGSKLIIKIST
RILSKRPATLESIODVTEIFALNRQRTAAGVFNSODGLVLDHNGLITWNPVTR
MLGPQEELIGKSLKAPDWOQIOTIMPTIOESLEHYCQWQGEVLEKHYNGNLPMFA
KVSRSKVGQVEQVDFVWIIISLDSGVKEMERLEYLAHHDALTGLANRKLHRTLEDE
LKLGNFTSEFALLYMDLDFKFINDFPGHDGDLVLRKVSERLNEVRSGLVARLS
GDEFVLIIRADNHSVTLISERILESRLQPIEYLNQTMQVGCISGVKLVGPQEKDME
ILKGADTAMYKAKSGKGAAILIGSEVGEACLSDVTE"
complement(2838..4703)
/locus_tag="VV20276"
complement(2838..4703)
/locus_tag="VV20276"
/note="similar to mostly Fe transport; COG1629"
/codon_start=1
/transl_table=11
/product="Outer membrane receptor protein"
/protein_id="AAO07241.1"
/db_xref="GI:27358291"
/translation="MEGRKRVKILVDGASQPCVFDGGPYSFINSASVSDPMLKSVEI
VKGAASSLHGSDAIGGVAFETKDPADFLKNGKVFQGAQLSYSEDSKFSFHVAVAK
RFDVNEALVATRRDGKELQNFSAKYDDYSVESQDVYKNDLLIKLSQLSGDHRLEP
LGEVYNQNDSDIASSSYNFNSDDTTKQNLGLKHLWVDTTITDSITSLTLVYGLS
ENGLTNPKEASAGVPRVPNGQKQKDYQVAENKLELTQDKLETHLYVYGLS
YKTSRISNTNREFNSDNPDKIYVYFDKAKEMSGFLPFQDEISVLNDKLVLTGVRY
YDFTSDSDTVDESFLKPSDSAFTRGLGATYKVAQEGTLFTQISQGFRAFPNELYIT
DYDFPHGVTNRPNPNLESKLSYELGYRHNTIASTSEIALYSDYDTDFISQTRGG
GLTEYTVNISEATIKGIEVNTLDNRILANPLGTMTRFVAAYTVGBDGSGLNSV
NPNMAGVANYDAPNTLGTGSIKUNYTAKKNSHINLDGNGGTANQVLPSTVIDL
TAYFKQDVTVRAGINLNTNEKYVLMNDIRGEKELSDLTQAERNYINISVKYEF"
5180..6094
/locus_tag="VV20277"
5180..6094
/locus_tag="VV20277"
/note="COG0583"
/codon_start=1
/transl_table=11
/product="Transcriptional regulator"
/protein_id="AAO07242.1"
/db_xref="GI:27358292"
/translation="MILITKGVEVQSPITLALHILDAIDRRGSFASAAANELDRAPS
LSYQIKLEQDLDIMIDFRSGHKANFTDAGLLILERGTILIAETKLVNDATLLANGW
BEDISIAVDGIIGHFHFSLVDALGNVSKTRVRLOEELIAGCWSGLADLLVCPK
LEVLPHDVKADTIGKMSMTVAAENHYVHKRSGEFNDTAREKYRIIAIDTAREQPAI
TWNIIQKQPLTVTNLAQKVDALVTGLIGTGLPTDIATPLIEAGLKEIEGAEQCHLE
ILLAWRNQMGAEAKSWCIIQLKKNKUK"
complement(6143..8821)
/locus_tag="VV20278"
complement(6143..8821)
/locus_tag="VV20278"
/note="Other COG candidate exists; similar to
acetyltransferase; COG0454"
/codon_start=1
/transl_table=11
/product="Histone acetyltransferase HPA2"
/protein_id="AAO07243.1"
/db_xref="GI:27358293"
/translation="MHKLOKLEKPTSAVIGASORHLSAGNIWMKNLQSGFGDTIIP
VTRPSVAGVLAYPSVTELPDVLPAICTSRERNVAFKQLAEIGTEFIVULAST
DYPHQDDQVADACLRARESNNRVLGPNLSGLILPMQKNFASFPVTPKPGFIATIS
QSAVCTTILDWANDKEIGFSAFVSGNGLDIDFDELLDYLDITSDSAIRRSGLRVN
DARFI GAARAAGRNRILVLKGRSILGRALNKGADLTDIVDSAIRRSGMLRVN
NTHLFAAVETLTHSVPLRGERLAIITNGAPALMAVDTLFERGGOLANLSSETIEML
SOHLPSWSHNNPIDIVGDKKRYVITNAILDSDCADAVLIMHSPASVSDSVETAQ
AIVDAIQSHPRHPRNLTNWSGEOTAKPARNILTKAGIPTYRTPESAVVAFHLVEY
RNRQKLMETPTTAEPUHVSQVNAKAWSEHLMDEIAQDTHQLGSLKLFHFDVL
PTWIASDTSEAVHIAEQIGYFVAVKLRSLAKSDVQVGNLNSQEVASADA1L
DRTKLYPSANIHGLLVQGVAKLAGGEEIRVKVITDKVFGFVILLGGGSEWSEARDA

```



```

gene      88..1707
/genes="mco-3"
/locus_tag="PP2925"
CDS       88..1707
/locus_tag="mco-3"
/genes="mco-3"
/locus_tag="PP2925"
/notes="similar to GB:LI19401, GB:D16391, GB:Z22957,
PID:307296, PID:312050, GB:LI19401, GB:D16391, GB:Z22957,
PID:307296, and PID:312050; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="malate:quinone oxidoreductase"
/protein_id="AAN68533.1"
/db_xref="GI:24984529"
/translation="MKITIKPLAMAGLALIGMLNAQAETRKVDVLLVGGIMSS
LAVMLNLPQSWMEVERLDKVAESSNGWNNCTGSHALAEELNYTPKDKILISK
AVINESFQVTRFLAWQVQGVKLPNPFINTPMSFWGDDNLYTPFKKRYEALQA
SPFKPMQSEDEHQIRKWVPLMMEGRDNQKLVATWTFIGDINFGELTRQYVGLQ
TRNFDLKUSSEVDITRNDGSMHVEYKLNKNGSAATDARFLFIGAGGAALPLIQ
SGIDEARNYGGFPVGGSFVTDNPAQAQRMKAKYGAATGAPPMSPVHLDRVLQDK
RLIFGPFATFSPKFLQKQSLDLLASTVHNAMPMVRVGFEDLVQVYLGVLQSD
DREALRTYFPAKEDNRLWQAGORVOIHKDEALGVLYLGTENVTSRDSGLAGL
LGSPGASTAAPMLVLANPKVKNKVASREWQAKIKQIIPSGIRLNDHPDQLEKWA
YTNVEVLQFPVSPQSPESIRPADSQGVASR"
/complement(1717..3045)
/locus_tag="PP2926"
/locus_tag="PP2926"
/notes="identified by match to PFAM protein family HMM
PF00984"
/codon_start=1
/transl_table=11
/product="UDP-Glucose dehydrogenase, putative"
/protein_id="AAN68534.1"
/db_xref="GI:24984530"
/translation="MKVTVGTCVGLTQAVCLAQVCHSVLCMDVDADEVASLSBCH
PIEPGLALLENKLCAGRLFTTDAANAANYARLOFIAVGPPPODAGADLKHFVAV
VDSILSHADGPKVINKSTVPVGTVHRIRAKIAQVADTSRFQVSNPFEFKGSAVD
CMRPRIIIGGAABAEVLLRELILPFNRNREKFMMDARSAEUTKAANCMATKI
SFINIANIAEHGLADIEMVRIGISDPRIQYDFTYAGCGFGSGFPKDLQALRSAAE
AEGFTQALRAEYVNEQOKGRLFISKIQRHYPGGLRGKVPALWGLSFKPNNTDIEAS
SRVLLALAAAGARVAQHDPOAMERTRRHVGPRADQLQVPCDKDALRGADALIVTEM
QDVRVNLDEVQQLADRVVFDGRNLYEPHMASAGLAYVYIGRGCKIAPA"
/3230..4336
/locus_tag="PP2927"
/locus_tag="PP2927"
/notes="identified by match to PFAM protein family HMM
PF02390"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAN68535.1"
/db_xref="GI:24984531"
/translation="MDEARLHDFMGKLVGDMGAATLANVILGDELGLYRAMADSOAV
TPDLAEKTCGERLIVREMLNQAASGVVMDKGRVLPFEQAMALALEDSAYVAGG
AAVVALFDKDKLVLAAMRGDGLWGDHPCVSGSTERFPFGVYRTLVADWLPALE
GVVAKLQAKADVDCGCGHASTLVNAQAPFASFTVGYDTHQPSITANAREAGLE
GVSGFOQASAKYFPGDHLVCFDCLHDMGDPVGAHAYHAKADGTMVLVEFYAE
DSLEGNLTVPGLRFLFYAASTFTICTNSLSQEVGLGLQAQGEARLARLRAVFEAGFSRFR
ATQTPFNLIIEARK"
/complement(4467..5711)
/locus_tag="PP2928"
/locus_tag="PP2928"
/notes="identified by match to PFAM protein family HMM
PF02896"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAN68536.1"
/db_xref="GI:24984532"
/translation="MKKNVLIIGAGVAKVVAHKAQHNDLGRITAIASRNISKQAI
IDSVKAGSLVADIDQAFSLNALDVERATKALISETESQIVNVGSAFLNMSVLRACI
DTGVAYLDTAHEEPGKICETPPWYGVYEWKHEECQAKNITALLGVSFGDQVNVSYA
KLAQQYFDSIDSIDLVNAGSHGKYPATNDEPFEINFRFTQGVSWQNSQWTSNTM
FEVXTDDLPVYQSNLYLTGKDEVHSISKNLNVNPFNMFSGEYHINFTVLRKLG
LLSQPVKTAEGLVFLKVKAVLPDPASLAPGYTKTICIGLVKGYKDGQPRBFVI
YVADLHEBAYAEKDSQGISYTAGVPPVAAALLVARGEWDAKRMVNVVEELPAEPFLKAL
DVMGLPTRVDEKGRDPMWABA"
/complement(5744..6841)
/genes="nspc"
/locus_tag="PP2929"
/locus_tag="PP2929"
/notes="similar to GB:M80629, and PID:180492; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="carboxynorspermidine decarboxylase"
/protein_id="AAN68537.1"
/db_xref="GI:24984533"
/translation="WIKTPYVYLDIKTLNNEKIAVYREHSGAKALLAKCFATWSV
FLMCOYMDGTTSSLPFLKLGROKFAGETHAYSVANADDEVEMLNCDKIINSIG
QLQPCASQSEKVRGLVNPVQYSSDYLLADPARPFSRLGEWDPAKIEQVTEQISGM
FHNCGDFGLFDKMLTHIEERFCHLLHKRVWSVSGGIHFTGEDYADFACRLKA
FSEKYGVLYLEPGEAAITNSASLEVTLDLYNGKHLAVVDSIEAHLDDLILYRLN
AKMAPNDGHTMYVCGKSLAGDIFGEYQDFRPLAIGDRLSFIDTAGYTMVKKWFG
LKMPISIVQLDGSVEVVERGFEDYVSSLS"
/complement(7027..7944)
/locus_tag="PP2930"
/locus_tag="PP2930"
/notes="similar to GP:57225, and SP:PI7324; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="L-serine dehydratase, putative"
/protein_id="AAN68538.1"
/db_xref="GI:24984535"
/translation="MTLHIHTPLIESPLSLAAGRNILWLKDALQPCGSFKLRGVGHA
CEVHARGARHFVSSGNGAGLAVAYAGKLGVPVTVVPIPTTTERAKQLHLEDAKV
VGGSCWCEANALATHPDNDATHPDDPLWAGHASLVDEVAAGLKPDADVLSV
VGGGLLGVVEGLAKNGVDVPLAVTEGAASLSHAAQAGHSVELEIRIASVATSLGA
KRVAQDALMCVQHPVSHLVSDRAALEACERFLDLHRVLVEPAGGAALAVAYDAKAL
GALRNVLVVCGGATLDEQIQAWLQQAD"
/complement(7941..8765)
/locus_tag="PP2931"
/locus_tag="PP2931"
/notes="similar to GP:5459226, and GP:5459226; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein TIGR00148"
/protein_id="AAN68539.1"
/db_xref="GI:24984536"
/translation="WLPSPRPARAGFRARRESAFSVRARSRSRSTICVKKWQK
LILSAMPILPSEERQLTLVLKAIQALGSAARANWEILLDLHDHPSHVAIVNGH
LISGVSFPIAAPEQDQKALMQAGADQGREPVVLPDPTTLKGLTILRSATAIER
DSSGVPASLCVNTDVTGLDAMPNQLQQLPLGPAIPAVNPEPAMELMSIIQAALOR
SQGRMNKQAKVEARYVMQREGLPIVKGVEKKAASALGVTRYTYIYNLEQLRGASQ"
/complement(8698..10401)
/locus_tag="PP2932"
/locus_tag="PP2932"
/notes="identified by match to PFAM protein family HMM
PF01425"
/codon_start=1
/transl_table=11
/product="amidase family protein"
/protein_id="AAN68540.1"
/db_xref="GI:24984537"
/translation="MIEVTEVSIABLRDALETGTATTAVELVKAYLARIDAYDGDATAT

```

ALNAVVRNEALKEAEASDARRAKQVLSPLDGIPIYTAQDSYLVKGLTAASGSPAFK  
DIVAORDAFTEIRLRAAGVCLCTNMPPMANGMORGCVTPNSYADNYLTAPFAS  
GSSNAGATAAASFAELAEWSSGRGPANNGICAVTPSRGIVSVRGNWPLPTM  
DVVPYARTWADLEILDVVVADADAKRGLRLOPWPPIPAASVTPASYLELAUDA  
SALKGRFFGVPMYINADPEAGTAEPKPGIGPTGQRINRATVIDLWQARALBAAG  
AEVLVDFFLVNCEGDEPGATVYRNGRIVSKAFDELWELSGWAFDFFLRANDPK  
LNLADVDGLKIFPHDPGLTPNREDLLAAGMEYVNNKRGKTNQDIELTPDGLRGL  
EQTKRLDEMDKLDGLDVLFTPTVADVPADADINPVSGADTAMSGVWVANGLAIR  
HLGVPTVTPVMGMADIMPVGLTFAGKAYSDNALLSFAAFAEATGSRMRWPPRTPAL  
G  
Complement(10593..11435)  
/locus\_tag="PP2933"  
complement(10593..11435)  
/locus\_tag="PP2933"  
/note="similar to GB:M59829, SP:P34931, SP:P48741, and  
2.8%; Score 99.4; DB 1; Length 300395;  
Best Local Similarity 55.1%; Pred. No. 1.9e-10;  
Matches 194; Conservative 0; Mismatches 137; Indels 3; Gaps 1;  
Qy 3252 AACATAAGCGCGTTAGTAAGAACCCTGTACGATCACCTATGACGTTGAACGAATGGCG 3311  
Db 171579 AACTCGACCGCGTTGCGCGCGCTCGGTCCATATCACCTACGACGTGGACATCGCGCATG 171520  
Qy 3312 CCGTAAGAGAAAGAGCTGCGCTTTGTTGTCGCTCATTCGCGACTTTTCAGGACACA 3371  
Db 171519 CCATCGAAGAAAGAGCTGCGCTTCGTCGTCGCGTACTTCGCGCAACC 171460  
Qy 3372 AACAGAGATCAGAAAAAGTTGATTAGAGAGCGAGAGTTCACGGGTATCGATAAAGACA 3431  
Db 171459 CGCGGAGCC---CCTGCCNAGCTCAAGGACCGCAAGTTCGTATCATCGATCGCGACA 171403  
Qy 3432 ACTTCGATACAGTGAATGGGCAATTACCGCGCTCTTTTCGTACAGGTTGATAACAGC 3491  
Db 171402 ACTTCAACGGTGTCTCAAGGGCATCAAGCCACGCTGACCTACCGCTCGACACACCC 171343  
Qy 3492 TTGCTATGATGATAGCGAGTTTCAAGTGAACCTTGACCTCCGTTGATCAAGAGATTTC 3551  
Db 171342 TGGCCAGACGGTACCGAGTTTGGCGGTGAGCTGAGTCAACAGCCTCGAAGACTTCG 171293  
Qy 3552 ACCCAGAGAACTTAGTTGATNAATTCAGCGCT 3585  
Db 171292 AGCCACAGAACGTTGTTCAAGCAGGTCGAGCGCT 171249  
RESULT 5  
AE016924/c  
LOCUS  
DEFINITION Chromobacterium violaceum ATCC 12472 section 15 of 16 of the complete genome.  
ACCESSION AE016924 AE016825  
VERSION AE016924.1 GI:34105231  
KEYWORDS Chromobacterium violaceum ATCC 12472  
SOURCE Chromobacterium violaceum ATCC 12472  
ORGANISM Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Chromobacterium.  
1 (bases 1 to 301457)  
REFERENCE  
AUTHORS Brazilian National Genome Project Consortium  
CONSTRM The complete genome sequence of Chromobacterium violaceum reveals remarkable and exploitable bacterial adaptability  
TITLE Proc. Natl. Acad. Sci. U.S.A. 100 (20), 11660-11665 (2003)  
JOURNAL 22822880  
MEDLINE 14500782  
PUBMED  
REFERENCE  
AUTHORS Vasconcelos, A.T.R., de Almeida, D.F., Almeida, F.C., de Almeida, L.G.P., de Almeida, R., Goncalves, J.A., Andrade, E.M., Antonio, R.V., Araripe, J., de Araujo, M.F., Filho, S.A., Azevedo, V., Batista, A.J., Bataus, L.A.M., Batista, J.S., Belo, A., vander Berg, C., Blamey, J., Bogo, M., Bonato, S., Bordignon, J., Brito, C.A., Brocchi, M., Burity, H.A., Camargo, A.A., Cardoso, D.D.P., Carneiro, N.P., Carraro, D.M., Carvalho, C.M.B., Cascardo, J.C.M.,

Cavada, B.S., Chueire, L.M.O., Pasa, T.B.C., Duran, N., Pagundes, N., Falcao, C.L., Fantinatti, F., Farias, I.P., Felipe, M.S.S., Ferrati, L.P., Ferro, J.A., Ferro, M.I.T., Franco, G.R., Freitas, N.S.A., Furlan, L.R., Gazzinelli, R.T., Gomes, E.A., Goncalves, P.R., Grangeiro, T.B., Grattapaglia, D., Grissard, E.C., Guimaraes, C.T., Hanna, E.S., Hungria, M. Jardim, S.N., Laurino, J., Leoi, L.C.F., Passarelli, L., Lima, A., Loureiro, M.F., Lyra, M.C.P., Macedo, M., Maderia, H.M.F., Manlio, G.P., Maranhao, A.Q., Martins, W.S., Di Mauro, S.M.Z., de Medeiros, S.R.B., Meissner, R.V., Menck, C.F.M., Moreira, M.A.M., Nascimento, F.F., Nicolas, M.F., Oliveira, J.G., Oliveira, S.C., Paixao, R.F.C., Parente, J.A., Pedrosa, F.O., Pena, S.J.D., Perreira, J.O., Perreira, M., Pinto, L.S.R.C., Pinto, L.S., Porto, J.I.R., Potrich, D.P., Neto, C.E.R., Reis, A.M.N., Rigo, L.U., Rindinelli, E., dos Santos, E.B.P., Santos, F.R., Schneider, M.P.C., Seanez, H.N., Silva, A.M.R., da Silva, A.L.C., Silva, D.W., Silveira, S., Simoes, I.C., Simon, D., Soares, C.M.A., Soares, R.B.A., Souza, E.M., Souza, K.R.D., Souza, R.C., Steffens, M.B.R., Steindel, M., Teixeira, S.R., Urmenyi, T., Vettore, A., Wassen, R., Zaha, A. and Simpson, A.J.G.  
Direct Submission  
Submitted (22-JAN-2003) Labinfo, LNCC - Laboratorio Nacional de Computacao Cientifica, Rua Getulio Vargas 333, Petropolis, RJ 25651070, Brazil  
Location/Qualifiers  
1. 301457  
/organism="Chromobacterium violaceum ATCC 12472"  
/mol\_type="genomic DNA"  
/strain="ATCC 12472"  
/db\_xref="taxon:243365"  
121..1386  
/gene="goac"  
/locus\_tag="CV3926"  
121..1386  
/gene="goac"  
/locus\_tag="CV3926"  
/EC\_number="2.6.1.19"  
/note="identified by sequence similarity; putative; ORF located using Glimmer/Blastx/COG0160"  
/codon\_start=1  
/trans\_table=1  
/product="4-aminobutyrate transaminase"  
/protein\_id="AAQ61587.1"  
/db\_xref="GI:34105232"  
/translation="MSQQQLNQRKSDATPRGVGVMCSFFPADRAKNAEWNTEGRYID FAGGIVLNTGHLHPKVEAAVAEQLSRFSHTQYQVYFYESYISAEKINALAPGRHAK KTAFTTGAEAVENAKIARATPGVIAFGGAFHGTLMGALTGKVAIKLGGFGP FPSDVHAPYPTNLHGVSQDSLQLEKLPKFDIPARVAIIYEPVQEGGFFAAPA EWKALPCLDCHGILLIADVEQSGFARTKLLASEHFDVAPDLITWAKSLAGGFPIL AVTGRAEVDAPAGGILGTYAGNPLALARAHAHLVDVDSKLAERQRLGDLKEKL NGLSAVPQIAEVRGLGMIAVEFNPQSGHQDPDFPAKQVQOHALANGLILLTCGIYA NVVRFLPFTTIEDIEIFAEALGKIEALKA"  
1456..2916  
/gene="gabb"  
/locus\_tag="CV3927"  
1456..2916  
/gene="gabb"  
/locus\_tag="CV3927"  
/EC\_number="1.2.1.16"  
/note="identified by sequence similarity; putative; ORF located using Glimmer/Blastx/COG1012"  
/codon\_start=1  
/trans\_table=1  
/product="succinate-semialdehyde dehydrogenase [NAD(P)]"  
/protein\_id="AAQ61588.1"  
/db\_xref="GI:34105233"  
/translation="MLNLQDPSLLRQOCLGGEWLDADGGQIHVHPATGEKTCQVP KMAEAEERAAVAARAFPAWKAKEKRAAIIIRWFDLIIHQDQDLIILTSQKRP LAEAGEIAYGASVIEWYAEKXVYDIPGAGDRVLVTKPEIGVTRAITPWNFP SAMITRKAAPLAAGCTWVRPASQTPLSALALAEARAGIPAGVSVITGSGSEIG AVLTGSDVTKPFSFTGVEGRKLLAOCAGTVKVMELGNAPPIVFDADLDAAVE KILSKYNNAGCTVCANRITYQGVYDPAFAEKFAAVALTKVGNLEAGVQOGLID HNAVAKVEEIRADALAGRLIAGKHALGHTTFEPILLADVTMDKMAVEEFGPL APLFFKTEEEAIRLANDTEFLGASYFYARDVGRIFRVAEGLEYMKNVNSGLLSNEA



Chandler, M., Choise, N., Claudel-Renard, C., Cunnac, S., Demance, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schlex, T., Siguier, P., Thebaud, P., Whalen, M., Wincker, P., Levy, M., Weissenbach, J. and Boucher, C.A.  
Genome sequence of the plant pathogen *Ralstonia solanacearum*  
Nature 415 (6871), 497-502 (2002)  
21681879  
PUBMED  
11823852  
2 (bases 1 to 190050)  
Boucher, C.A.  
Direct Submission  
Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie Molculaire des Interactions Plantes-Microorganismes INRA-CNRS, BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS 118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27, F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire INRA, BP27, F31326 Castanet-Tolosan Cedex  
Christian.Boucher@oulouse.inra.fr  
http://sequence.toulouse.inra.fr/R.solanacearum.html.  
FEATURES  
source  
1..190050  
Location/Qualifiers  
/organism="Ralstonia solanacearum"  
/mol\_type="Genomic DNA"  
/strain="GMI1000"  
/db\_xref="taxon:305"  
/plasmid="megaplasmid"  
51..362  
/gene="RSp0643"  
/note="synonym: RS05579"  
51..362  
/gene="RSp0643"  
/functions="miscellaneous; unknown"  
/note="Product confidence : hypothetical  
Gene name confidence : hypothetical  
predicted by Codon usage  
predicted by FrameD"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="HYPOTHETICAL PROTEIN"  
/protein\_id="CAD17794.1"  
/db\_xref="GI:17431114"  
/db\_xref="GOA:Q8XS38"  
/db\_xref="SPTREMBL:Q8XS38"  
/translation="MKTSPSPNPAAGPASPDAAGDQVAFQOLHEYNHTITIRPR  
RIGGLWRAFAVSLDSALVAVGVTEVGOAFRTPIGARDVADARSFIDARLKACRP  
A"  
397..555  
/gene="RSp0644"  
/note="synonym: RS05578"  
397..555  
/gene="RSp0644"  
/functions="miscellaneous; unknown"  
/note="Product confidence : putative  
Gene name confidence : hypothetical  
predicted by Codon usage  
predicted by FrameD"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="PUTATIVE SIGNAL PEPTIDE PROTEIN"  
/protein\_id="CAD17795.1"  
/db\_xref="GI:17431115"  
/db\_xref="GOA:Q8XS37"  
/db\_xref="SPTREMBL:Q8XS37"  
/translation="MEKQLFAIAIVIAAASVSLARDLNDDERDPPQQAALVEGH  
ATASAAGR"  
Complement (575..1135)  
/gene="RSp0645"

CDS  
/note="synonym: RS05577"  
complement (575..1135)  
/gene="RSp0645"  
/function="miscellaneous; not classified regulator"  
/note="Product confidence : probable  
Gene name confidence : hypothetical  
predicted by Codon usage  
predicted by Homology  
predicted by FrameD"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="PROBABLE TRANSCRIPTION REGULATOR PROTEIN"  
/protein\_id="CAD17796.1"  
/db\_xref="GI:17431116"  
/db\_xref="GOA:Q8XS36"  
/db\_xref="SPTREMBL:Q8XS36"  
/translation="MKRISFPDPDMKVAIKERSPAAAPPOGAALSALDRDTRAIL  
KWLQDASISNVALADKVSPPACLRVERLKEGICRGVALLERKALDGLTLMVI  
GVLDRTSPESFADFKAQVKVPCGLECHVVTGDFYMLVRITKNDNSYNRLHAEQL  
YLPGRQIRITFTMLKQVLSITQLPIA"  
1244..2260  
/gene="RSp0646"  
/note="synonym: RS05576"  
1244..2260  
/gene="RSp0646"  
/EC\_number="4.1.99.4"  
/function="miscellaneous; hypothetical/global homology"  
/note="Product confidence : probable  
Gene name confidence : hypothetical  
predicted by Codon usage  
predicted by Homology  
predicted by FrameD"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="PROBABLE 1-AMINOCYCLOPROPANE-1-CARBOXYLATE  
DEAMINASE (ACC DEAMINASE) PROTEIN"  
/protein\_id="CAD17797.1"  
/db\_xref="GI:17431117"  
/db\_xref="GOA:Q8XS35"  
/db\_xref="SPTREMBL:Q8XS35"  
/translation="MNLKHPHPLTFTPTIQLKLSAHLGKGVLYAKREDNSG  
LAFGNKTRKLEYLVPEVLGGYDTLSIGIQSNQTRQVAAVAAHLGKLVQENW  
VNYADAVRVGNIELSLILCADVLDLDAAGFDIGIRPSWEQAMEDVRAGKPPPIPA  
CQSEHPICGLGFVGFABEVQQAEGFRFDYVVCSTGTQAGVVGFAADGRAD  
VIGIDASAKPEQTRQIRIARDIAKVELGRDITEDDVLDTYRGGEVGLNEGTL  
EAILRCASQEGMLTDPVYEGKSMHMDIRVRGGEFFEGSRVLYAHLGVPALNAYSFL  
FRNG"  
complement (2336..3508)  
/gene="RSp0647"  
/note="synonym: RS05575"  
complement (2336..3508)  
/gene="RSp0647"  
/EC\_number="4.2.1.17"  
/function="miscellaneous; hypothetical/global homology"  
/note="Product confidence : putative  
Gene name confidence : hypothetical  
predicted by Codon usage  
predicted by Homology  
predicted by FrameD"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="PUTATIVE ENOYL-COENZYME A HYDRATASE PROTEIN"  
/protein\_id="CAD17798.1"  
/db\_xref="GI:17431118"  
/db\_xref="GOA:Q8XS34"  
/db\_xref="SPTREMBL:Q8XS34"  
/translation="MSALPKPAEPAAABEVLPHVNVNVAILLNRPALNALSHGM  
VRQSLAALQCRDDPRTIAAVVLRGAGEKFGCAGGVDRALYELARQGAHQHGDWMLQFF  
VDEYELDALHRFPKPVALLDGIAMGGMLGQGARLVVTRSRITAMPETRIIGFLP  
DVGATHEFLGAMPPEMALVIGLTGTTLSCADALHCLGLADVCVPAEWLASFEDRLRLSLH

CDS

gene

CDS

gene

CDS

gene







```

Db      2074  ACTTCAGAGCTCATGCGCGCATGACGCGCGCCTCACCCCTGTCGGTCGAGAACACTC 2133
QY      3492  TTGCTAAATGATGATGACGAGTTTGAAGTGAACCTTGAGCCTCGTTTCGATGAAGATTTCC 3551
Db      2134  TCGAGAAAGCGACAGAGCTTCCGTTTCGCTGAATTCGAAAGATGCCGATTCG 2193
QY      3552  ACCCAGAGAACTTAGTTGATNAATTTGACCGCTTAA 3598
Db      2194  AACCCGCGCGTCTCGATCAGTGCAGGCGGCTGAA 2230

RESULT 8
AE001398
LOCUS   Plasmodium falciparum 3D7 chromosome 2 section 35 of 73 of the
DEFINITION complete sequence.
ACCESSION AE001398 AE001362
VERSION   AE001398.1 GI:3845197
KEYWORDS
SOURCE    Plasmodium falciparum 3D7
ORGANISM  Plasmodium falciparum 3D7
REFERENCE 1 (bases 1 to 14867)
AUTHORS   Gardner,M.J., Tettelin,H., Carucci,D.J., Cummings,L.M., Aravind,L.,
          Koonin,E.V., Shallem,S., Mason,T., Yu,K., Fujii,C., Pederson,U.,
          Shen,K., Jing,J., Aston,C., Lai,Z., Schwartz,D.C., Pertea,M.,
          Salzberg,S., Zhou,L., Sutton,G.G., Clayton,R., White,O.,
          Smith,H.O., Fraser,C.M., Adams,M.D., Venter,J.C. and Hoffman,S.L.
          Chromosome 2 sequence of the human malaria parasite Plasmodium
          falciparum
          Science 282 (5391), 1126-1132 (1998)
JOURNAL   Science
MEDLINE   99021243
PUBMED   9804551
REFERENCE 2 (bases 1 to 14867)
AUTHORS   Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W.,
          Carlton,J.M., Pain,A., Nelson,K.E., Bowman,S., Paulsen,I.T.,
          James,K., Eisen,J.A., Rutherford,K., Salzberg,S.L., Craig,A.,
          Kyes,S., Chan,M.-S., Nene,V., Shallem,S.J., Suh,B., Peterson,J.,
          Agiuoli,S., Pertea,M., Allen,J., Selengut,J., Haft,D.,
          Mauchner,M.W., Vaidya,A.B., Martin,D.M.A., Fairlamb,A.H.,
          Fraunholz,M.J., Roos,D.S., Ralph,S.A., McFadden,G.I.,
          Cummings,L.M., Subramanian,G.M., Mungall,C., Venter,J.C.,
          Carucci,D.J., Hoffman,S.L., Newbold,C., Davis,R.W., Fraser,C.M. and
          Barrell,B.
          Genome sequence of the human malaria parasite Plasmodium falciparum
          Nature 419, 498-511 (2002)
JOURNAL   Nature
MEDLINE   12021433
PUBMED   12021433
REFERENCE 3 (bases 1 to 14867)
AUTHORS   Gardner,M.J.
          Direct Submission
          Submitted (02-OCT-2002) The Institute for Genomic Research, 9712
          Medical Center Dr, Rockville, MD 20850, USA
FEATURES
         source
         1..14867
         /organism="Plasmodium falciparum 3D7"
         /mol_type="genomic DNA"
         /isolate="3D7"
         /db_xref="taxon:36329"
         /chromosome="2"
         complement(1..20)
         /note="AT rich"
         /rpt_type=tandem
         complement(65..137)
         /note="AT rich"
         /rpt_type=tandem
         complement(178..211)
         /rpt_type=tandem
         /rpt_unit="{TA}n"
         307..327
         /rpt_type=tandem
         /rpt_unit="{A}n"
         complement(328..348)
         /rpt_type=tandem
         /rpt_unit="{TA}n"

repeat_region complement(378..423)
                /note="AT rich"
                /rpt_type=tandem
repeat_region complement(479..503)
                /note="AT rich"
                /rpt_type=tandem
repeat_region complement(513..539)
                /rpt_type=tandem
                /rpt_unit="{A}n"
                577..633
                /rpt_type=tandem
                /rpt_unit="{A}n"
                complement(714..747)
                /note="AT rich"
                /rpt_type=tandem
                841..867
                /rpt_type=tandem
                /rpt_unit="{A}n"
                complement(888..918)
                /note="AT rich"
                /rpt_type=tandem
                complement(1006..1049)
                /note="AT rich"
                /rpt_type=tandem
                complement(1075..1120)
                /note="AT rich"
                /rpt_type=tandem
                complement(1140..1283)
                /note="AT rich"
                /rpt_type=tandem
                1294..1346
                /rpt_type=tandem
                1298..1346
                /rpt_type=tandem
                /rpt_unit="{TATAA}n"
                complement(1423..1531)
                /note="AT rich"
                /rpt_type=tandem
                complement(<1570..>2424)
                /locus_tag="PF0490c"
                complement(<1570..>2424)
                /locus_tag="PF0490c"
                complement(1570..2424)
                /locus_tag="PF0490c"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAC71887.1"
                /db_xref="GI:3845198"
                /translation="MKENEKIMDYLSQPLDDVDREKKSGKNSLLKSSSTKKSDDYK
                SSIFSKRDSHKSGSFRGRSGFNRKSGSKPKFYNNRLINKNNYKGNFNGR
                DNFKGTGSGSRVFDNRKSGPKKFIENRKSSVKSYGNSKMGKSKNPKPTSR
                TVVTXLNNYKTVSAPVKKFNINLSLYRKORTPALNTRKSPVGTIKSSVPKRIKK
                GLKGSLSKTRKSTSGSKFPLFKSLKIVTSLNKLPSPLKEQKNTEVNDPEIS
                NNAITTKN"
                complement(2150..2188)
                /note="AT rich"
                /rpt_type=tandem
                complement(2446..2468)
                /rpt_type=tandem
                2504..2656
                /rpt_type=tandem
                /rpt_unit="{TA}n"
                /rpt_type=tandem
                /rpt_unit="{TATAA}n"
                2684..2739
                /rpt_type=tandem
                2693..2736
                /rpt_type=tandem
                complement(2697..2732)
                /rpt_type=tandem
                /rpt_unit="{TA}n"
                complement(2757..2845)
                /note="AT rich"
                /rpt_type=tandem

```



```
REFERENCE 2 (bases 1 to 252394)
AUTHORS Gardner M.J.
TITLE Direct Submission
JOURNAL Submitted (13-SEP-2002) The Institute for Genomic Research, 9712
Location/Qualifiers
FEATURES
    source
        1. .252394
            /organism="Plasmodium falciparum 3D7"
            /mol_type="genomic DNA"
            /isolate="3D7"
            /db_xref="taxon:36329"
            /chromosome="10"
            complement(79. .213)
            /note="AT rich"
            /rpt_type=tandem
            253. .301
            /rpt_type=tandem
            complement(253. .301)
            /rpt_type=tandem
            /rpt_unit="(TA)n"
            318. .374
            /rpt_type=tandem
            /rpt_type=tandem
            324. .374
            /rpt_type=tandem
            /rpt_unit="(TAAA)n"
            complement(433. .482)
            /rpt_type=tandem
            /rpt_unit="(TA)n"
            577. .605
            /rpt_type=tandem
            complement(577. .605)
            /rpt_type=tandem
            /rpt_unit="(A)n"
            complement(661. .682)
            /rpt_type=tandem
            /rpt_unit="(TA)n"
            complement(792. .942)
            /rpt_type=tandem
            /rpt_type=tandem
            897. .1030
            /rpt_type=tandem
            /rpt_unit="(TATAA)n"
            complement(1038. .1173)
            /note="AT rich"
            1100. .1130
            /rpt_type=tandem
            /rpt_type=tandem
            complement(1100. .1126)
            /rpt_type=tandem
            /rpt_unit="(TA)n"
            1133. .1172
            /rpt_type=tandem
            1175. .1207
            /rpt_type=tandem
            complement(1175. .1207)
            /rpt_type=tandem
            /rpt_unit="(A)n"
            complement(1218. .1281)
            /note="AT rich"
            1365. .1391
            /rpt_type=tandem
            /rpt_type=tandem
            complement(1365. .1391)
            /rpt_type=tandem
            /rpt_unit="(TA)n"
            complement(1401. .1493)
            /note="AT rich"
            1505. .1526
            /rpt_type=tandem
            /rpt_type=tandem
            complement(1602. .1685)
            /rpt_type=tandem
            /rpt_unit="(A)n"
            repeat_region
                /rpt_unit="(TA)n"
                complement(1726. .1825)
                /note="AT rich"
                /rpt_type=tandem
                repeat_region
                    /note="AT rich"
                    complement(1963. .2007)
                    /note="AT rich"
                    /rpt_type=tandem
                    complement(2049. .2157)
                    /rpt_type=tandem
                    /rpt_unit="(TA)n"
                    2064. .2131
                    /rpt_type=tandem
                    complement(2194. .2390)
                    /rpt_type=tandem
                    /rpt_unit="(TA)n"
                    2257. .2332
                    /rpt_type=tandem
                    complement(2405. .2457)
                    /note="AT rich"
                    /rpt_type=tandem
                    complement(2481. .2516)
                    /note="AT rich"
                    /rpt_type=tandem
                    2532. .2615
                    /rpt_type=tandem
                    /rpt_type=tandem
                    /rpt_unit="(A)n"
                    complement(2623. .2657)
                    /note="AT rich"
                    /rpt_type=tandem
                    2666. .2699
                    /rpt_type=tandem
                    /rpt_type=tandem
                    /rpt_unit="(TAAA)n"
                    complement(2740. .2807)
                    /note="AT rich"
                    /rpt_type=tandem
                    2855. .3047
                    /rpt_type=tandem
                    /rpt_type=tandem
                    complement(2856. .3047)
                    /rpt_type=tandem
                    /rpt_type=tandem
                    /rpt_unit="(TA)n"
                    2893. .3047
                    /rpt_type=tandem
                    /rpt_type=tandem
                    3100. .3137
                    /rpt_type=tandem
                    complement(3100. .3137)
                    /rpt_type=tandem
                    /rpt_type=tandem
                    /rpt_unit="(TA)n"
                    complement(3145. .3170)
                    /note="AT rich"
                    /rpt_type=tandem
                    complement(3196. .3240)
                    /note="AT rich"
                    /rpt_type=tandem
                    3283. .3319
                    /rpt_type=tandem
                    complement(3287. .3319)
                    /rpt_type=tandem
                    /rpt_unit="(TA)n"
                    complement(3367. .3434)
                    /note="AT rich"
                    /rpt_type=tandem
                    complement(3513. .3656)
                    /rpt_type=tandem
                    /rpt_unit="(TA)n"
                    complement(3659. .3794)
                    /rpt_type=tandem
                    /rpt_unit="(TA)n"
                    complement(3809. .3838)
                    /note="AT rich"
                    /rpt_type=tandem
                    3840. .3861
                    /rpt_type=tandem
                    /rpt_unit="(A)n"
```

repeat_region	complement(3866..4026)	Db	147542	TATATCTATAATAATAATAA	147561
repeat_region	/note="AT rich"				
repeat_region	/rpt_type=tandem				
repeat_region	complement(4089..4110)				
repeat_region	/note="AT rich"				
repeat_region	4111..4295				
repeat_region	/rpt_type=tandem				
repeat_region	complement(4111..4235)				
repeat_region	/rpt_type=tandem				
repeat_region	/rpt_unit="(TA)n"				
repeat_region	complement(4251..4414)				
repeat_region	/note="AT rich"				
repeat_region	/rpt_type=tandem				
repeat_region	complement(4262..4295)				
repeat_region	/rpt_type=tandem				
repeat_region	/rpt_unit="(TA)n"				
repeat_region	complement(4361..4384)				
repeat_region	/rpt_type=tandem				
repeat_region	complement(4424..4483)				
repeat_region	/rpt_type=tandem				
repeat_region	/rpt_unit="(TA)n"				
repeat_region	complement(4511..4553)				
repeat_region	/rpt_type=tandem				
repeat_region	/rpt_unit="(TA)n"				
repeat_region	complement(4602..4628)				
repeat_region	/note="AT rich"				
repeat_region	/rpt_type=tandem				
repeat_region	complement(4663..4725)				
repeat_region	/note="AT rich"				
Query Match 2.1%; Score 76.8; DB 3; Length 25394;					
Best Local Similarity 48.4%; Pred. No. 1.1e-05;					
Matches 271; Conservative 0; Mismatches 287; Indels 2; Gaps 2;					
QY	541 ATATTCATTAATTCACATTAATAAACAAGCTCTAAAGGCTGTTTATTGGATGAATAT				600
Db	147002 AAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA				147061
QY	601 TCGAATATACATATAAATGATGCTATTATT-ACCTGCTGATGGATCAACTTC				659
Db	147062 ATAAATATTAATAAAGATTAGTTTAAATTAATTAATTTATTTTATTTATATAT				147121
QY	660 ATGCTCTATACATGA-ATATATTTTCAGTTAGACCTTAATTCAGGTAATTCCTATT				718
Db	147122 TAT				147181
QY	719 TAAATTAATTCGATAATATGTAATCGATTGCTTTGTTGGTTATTTTATGTTTCTCA				778
Db	147182 TTAATTTTATATATATTTCAATTTTTTTTTTTTTTTTGTGTTTTCGTTCTTAAT				147241
QY	779 TTTTAAATGACGGTGAGCTGGCGATTCATATTTTATGATGACACAATCTTTGATCAA				838
Db	147242 ATTATATATTAAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATA				147301
QY	839 GTATTTAAGATATTCGTAATGATGAGGGTTTGGCTGATTTTATTTATATAAATCAATA				898
Db	147302 ATTTTAAATGACTGTAAATAATACATTAATTCATAAATAATGATAATTTATATAT				147361
QY	899 TAAATACAAATATATGTTATTTTGTGCTCTTTTATAGTGTCTTTTAAAGAGGTAGGA				958
Db	147362 ATAAATGAAGTTTCTTTTTTTTTTTTTTTTTTTTGGGTATATATATATATATAT				147421
QY	959 TGACCTAAGGTCGCTAAATATGCGTAATTCGCAATTCGTAATTCACCTCAAGAT				1018
Db	147422 AT				147481
QY	1019 ACACATTTGGCAATGACAAATATGTCACCTCGATGAAACAATATAGTAGATGTTGT				1078
Db	147482 ATATAAATATGTAATATACATATATAAATAATATATATAAATAAATAAATAAATA				147541
QY	1079 TTTTCTGCAAAAATAAAAA				1098
RESULT 10					
AP003630/1c					
LOCUS					
DEFINITION					
Oryza sativa (japonica cultivar-group) chromosome 6 clone P0566A10,					
*** SEQUENCING IN PROGRESS ***.					
ACCESSION					
AP003630.1 GI:14141691					
VERSION					
HTG, HTGS, PHASE2					
KEYWORDS					
Oryza sativa (japonica cultivar-group)					
SOURCE					
ORGANISM					
Oryza sativa (japonica cultivar-group)					
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;					
Ehrhartoideae; Oryzeae; Oryza.					
REFERENCE					
1					
Sasaki, T., Matsumoto, T. and Yamamoto, K.					
Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC					
clone:P0566A10					
Published Only in Database (2001)					
2 (bases 1 to 179902)					
Sasaki, T., Matsumoto, T. and Yamamoto, K.					
Direct Submission					
Submitted (17-MAY-2001) Takuji Sasaki, National Institute of					
Agrobiological Resources, Rice Genome Research Program; Kannondai					
2-1-2, Tsukuba, Ibaraki 305-8602, Japan					
(E-mail:tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,					
Tel:81-298-38-7441, Fax:81-298-38-7468)					
NOTE: It currently consists of 1 contigs. Gaps between the contigs					
are represented as runs of N. The order of the pieces is believed					
to be correct as given, however the sizes of the gaps between them					
are based on estimates that have provided by the submitter. This					
sequence will be replaced by the finished sequence as soon as it is					
available and the accession number will be preserved.					
* NOTE: This is a 'working draft' sequence.					
* This sequence will be replaced					
* by the finished sequence as soon as it is available and					
* the accession number will be preserved.					
FEATURES					
source					
1..179902					
/organism="Oryza sativa (japonica cultivar-group)"					
/mol_type="genomic DNA"					
/cultivar="Nipponbare"					
/db_xref="taxon:39947"					
/chromosome="6"					
/clone="P0566A10"					
ORIGIN					
Query Match 2.1%; Score 76.2; DB 2; Length 179902;					
Best Local Similarity 48.5%; Pred. No. 1.5e-05;					
Matches 210; Conservative 0; Mismatches 223; Indels 0; Gaps 0;					
QY	514 GTTTGGTGTCTAAATCGTCTACTGCTCATATTCATTAATTCACATTAATAAATAACAGTT				573
Db	63539 GTTTTGTCTGTATGTTTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT				63480
QY	574 CTAAGAGGCTGTTTATGATGAATATTCGAAATATACATAAATAATGATGCTATATAT				633
Db	63479 TTTTGTCTGTATGTTTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT				63420
QY	634 TACTGCTGTATGTTTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT				693
Db	63419 TGTCTGTATGTTTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT				63360
QY	694 CTTAATTCAGGTAAATTCGCTATTAATTAATTAATTAATTAATTAATTAATTAATTCG				753
Db	63359 TTTTGTCTGTATGTTTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT				63300
QY	754 TGTGGTATTTTATGTTTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT				813
Db	63299 TTTTGTCTGTATGTTTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT				63240



repeat_region	26815..27110	/note="AluSq repeat: matches 1..303 of consensus"
repeat_region	27111..27285	/note="Aluwb repeat: matches 130..307 of consensus"
repeat_region	27286..27578	/note="L2 repeat: matches 2359..2668 of consensus"
repeat_region	29945..30257	/note="AluY repeat: matches 1..306 of consensus"
repeat_region	30812..31078	/note="AluX repeat: matches 1..267 of consensus"
repeat_region	31887..32091	/note="MER58A repeat: matches 1..222 of consensus"
repeat_region	32818..32853	/note="9 copies 4 mer acac 100% conserved"
repeat_region	32856..32957	/note="MER97C repeat: matches 185..289 of consensus"
repeat_region	32958..33406	/note="MLRIC repeat: matches 1..466 of consensus"
repeat_region	33407..33867	/note="MER97C repeat: matches 289..726 of consensus"
repeat_region	33895..34718	/note="8 copies 103 mer 61% conserved"
repeat_region	33915..34519	/note="121 copies 5 mer atata 56% conserved"
repeat_region	33920..34103	/note="92 copies 2 mer at 57% conserved"
repeat_region	33924..34553	/note="21 copies 30 mer 61% conserved"
repeat_region	33925..34730	/note="13 copies 62 mer 61% conserved"
repeat_region	33940..34709	/note="14 copies 55 mer 61% conserved"
repeat_region	33954..34213	/note="13 copies 20 mer 70% conserved"
repeat_region	33999..34698	/note="25 copies 28 mer 58% conserved"
repeat_region	34014..34650	/note="13 copies 49 mer 62% conserved"
repeat_region	34071..34642	/note="11 copies 52 mer 62% conserved"
repeat_region	34300..34709	/note="205 copies 2 mer aa 56% conserved"
repeat_region	34390..34689	/note="15 copies 20 mer 60% conserved"
repeat_region	34488..34699	/note="53 copies 4 mer taaa 59% conserved"
repeat_region	34512..34703	/note="64 copies 3 mer ata 60% conserved"
repeat_region	34629..34718	/note="3 copies 30 mer 78% conserved"
repeat_region	35082..35647	/note="LiMEI repeat: matches 5550..6144 of consensus"
repeat_region	35694..35885	/note="L2 repeat: matches 2054..2250 of consensus"
repeat_region	35882..36229	/note="L2 repeat: matches 1446..1832 of consensus"
repeat_region	36377..36484	/note="MIR repeat: matches 48..147 of consensus"
repeat_region	39838..39877	/note="20 copies 2 mer tt 82% conserved"
repeat_region	39841..39875	/note="7 copies 5 mer tggtt 85% conserved"
repeat_region	41466..41514	/note="L2 repeat: matches 2699..2747 of consensus"
repeat_region	41792..42143	/note="RH31B repeat: matches 1..364 of consensus"
repeat_region	43471..43590	/note="L2 repeat: matches 2629..2750 of consensus"
repeat_region	44937..45157	/note="L2 repeat: matches 2259..2489 of consensus"
repeat_region		
Query Match	2.1%;	Score 75.4; DB 9; Length 143409;
Best Local Similarity	47.6%;	Pred. No. 2.3e-05;
Matches 313; Conservative 0; Mismatches 341; Indels 3; Gaps 3;		
QY	461	ATTGCGTGACATAAAGCTTTTCTATCTGAAAGCTTAATGCTGAGGAGCGGTTTGGT 520
Db	34685	AITTTTATATAATTTTAT 34686
QY	521	GCTTTAAATTCGTTACTGTCATATTCATTAATTCACATTAATTAATAAAGCTCTAAAG 580
Db	34625	TATTTATATAATATATTTTATATAATATATATATATATATATATATATATATATATAT 34566
QY	581	GCTGTTTATTCGATGAATATTCGAATATACATAAATTAATTCGATTAATTAATTTACTTC 640
Db	34565	AATATATATTTTA 34506
QY	641	TGTATTTGGTATCAACTTTTCATGCTCTATACATGTAAATATATTTCCGAGTTAGACCTTAAT 700
Db	34505	TATATTTATATTA-TATAATA 34447
QY	701	CAAGGTAATTTGCTCTATTTAATTTATATCTGAATATATGTAATCGATTCGTTTCTGGTT 760
Db	34446	TTATATAATATATATTTTAT 34387
QY	761	ATTTT-TATGTTTGTTCATTTTAAATGACGGTGAGCTTGTGCATTCATATTTTTTATGA 819
Db	34386	ATTATATATTTTATATAATA 34327
QY	820	TGACAACTCTTTGATGAAGTATTTAAGATATTTGTAATGATGAGGGGTTTGGCTGTAT 879
Db	34326	TTTATATTTTATTTTAT 34267
QY	880	TTTTTATATTAATCATATAATAAATCAACAATATATGTTATTTTGTCTTTTATATAGT 939
Db	34266	ATTATATATTAATATATGTTATATATAACATATATATATATATATATATATATATATAT 34207
QY	940	TTCTTTTAAAGAGTAGGATGACCTTAAGGTCGCCCTAAATATGCGGTAATTTGCCAT-TG 998
Db	34206	ATATTATGTTTAAAT 34147
QY	999	CTATATATTCACCTCAAGATACACTATTGGCAAAATGACAAATATGTCATTCGTATGAA 1058
Db	34146	ATATATTTGTTATTTAAAT 34087
QY	1059	ACAATATTAGTAGATGTTGTTTTTCTGCGAAATAAATAATTTTCTGGTGAATA 1115
Db	34086	ATTATATATGAATGA 34030
RESULT 12		
EX005205		
LOCUS		
DEFINITION	164188 bp DNA linear HTG 30-OCT-2003	
ACCESSION	EX005205	
VERSION	EX005205.10	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.	
SOURCE	Danio rerio (zebrafish)	
ORGANISM	Danio rerio	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Actinopterygii; Neopterygii; Teleostei; Ostariophysi;	
TITLE	Cypriniformes; Cyprinidae; Danio.	
JOURNAL	1 (bases 1 to 164188)	
COMMENT	Direct Submission	
	Submitted (29-OCT-2003) Wellcome Trust Sanger Institute, Hinxton,	
	Cambridgeshire, CB10 1SA, UK. E-mail enquiries:	
	zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk	
	On Sep 27, 2003 this sequence version replaced gi:33468698.	
	----- Genome Center	
	Center: Wellcome Trust Sanger Institute	
	Center code: SC	
	Web site: http://www.sanger.ac.uk	
	Contact: zfish-help@sanger.ac.uk	
	----- Project Information	

```

Center project name: zK210P14
----- Summary Statistics -----
Assembly program: XGAF4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 163185 bases at least Q40
Consensus quality: 163490 bases at least Q30
Consensus quality: 163732 bases at least Q20
Insert size: 163888; sum-of-contigs
Quality coverage: 6.58x in Q20 bases; sum-of-contigs Quality
coverage: 6.92x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 82331: contig of 82331 bp in length
* 82432: gap of 100 bp
* 82432: 85590: contig of 3159 bp in length
* 85591: 85690: gap of 100 bp
* 85691: 107541: contig of 21851 bp in length
* 107542: 107642: gap of 100 bp
* 107642: 164188: contig of 56547 bp in length.
*
* Location/Qualifiers
* 1..164188
* /organism="Danio rerio"
* /mol_type="genomic DNA"
* /db_xref="taxon:7955"
* /clone="DKEY-210P14"
* /clone_lib="DanioKey"
*
* misc_feature
* 1..82331
* /note="assembly fragment: 01879"
* fragment_chain:1
*
* misc_feature
* 82432..85590
* /note="assembly fragment: 00792"
* fragment_chain:1
*
* misc_feature
* 85691..107541
* /note="assembly fragment: 01113"
* fragment_chain:1
*
* misc_feature
* 107642..164188
* /note="assembly fragment: 01413"
* fragment_chain:1
*
ORIGIN
Query Match 2.1%; Score 74.8; DB 2; Length 164188;
Best Local Similarity 51.3%; Pred. No. 3e-05;
Matches 199; Conservative 0; Mismatches 187; Indels 2; Gaps 1;

QY 558 TATAATAAACAGTTCTTAAAGGCTGTTTATTGGATGAATATCGAAATTATCATATAA 617
DB 135066 TTTTAAATATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 135125

QY 618 TAATTCATGCTATTATCTGCTGATTGGTATCAACTTCATGCTCTATACATGTAAT 577
DB 135126 TTAATTTATTATTATTGTTTATTATTGACTTTTATTATTATTATTATTATTATTATT 135185

QY 678 ATATTTTCGAGTTAGACCTTAATCAAGGTAATTTGCTATTTAATTAATTTATCTGAATAAT 737
DB 135186 TTAATTCATTTATTATTATTTTCAATTTTAGTTGTTGTTTATTATTGACTATTATTGTAATTT 135245

QY 738 ATGTAATCATGCTTTGTTGGTTATTTTATGTTTGTTCATTTTAAATGACGGTGAGCT 797
DB 135246 TATTTTAAATAAAT- -TTTTTATTGTTTCTACTGTTTATTTTCAATTTGCTATTGCTATTATTAT 135303

QY 798 TGRGCAATTCATATTTTTTTATGATGACACATCTTTTGATCAAGTATTTTAAGATATTGTTAA 857
DB 135304 TTAATTTTGTGTTTATTATTATTGATTATTATTATTATTATTATTATTATTATTATTAA 135363

QY 858 TGCATGAGGGGTTTGGCTGTATTTTATTATATTAAATCAATAATAAATAACAATATATGT 917

```

```

DB 135364 AATCATTTGAGTATTTGTTTATTATTATTATTATTATTATTATTATTATTATT 135423
QY 918 TATTTTGTGCTCTTTTATAGTCTTTT 945
DB 135424 TATTTGTGATTTATTATTATTATTATT 135451

RESULT 13
PFMAL7P1_07
WPCOMMENT
Sequence split into 14 fragments LOCUS PFMAL7P1 Accession AL844506
Fragment Name Begin End
PFMAL7P1_00 1 110000
PFMAL7P1_01 100001 210000
PFMAL7P1_02 200001 310000
PFMAL7P1_03 300001 410000
PFMAL7P1_04 400001 510000
PFMAL7P1_05 500001 610000
PFMAL7P1_06 600001 710000
PFMAL7P1_07 700001 810000
PFMAL7P1_08 800001 910000
PFMAL7P1_09 900001 1010000
PFMAL7P1_10 1000001 1110000
PFMAL7P1_11 1100001 1210000
PFMAL7P1_12 1200001 1310000
PFMAL7P1_13 1300001 1351552
Continuation (8 of 14) of PFMAL7P1 from base 700001 (AL844506 Plasmodium falciparum 3D7)

```

```

Query Match 2.1%; Score 74.6; DB 2; Length 110000;
Best Local Similarity 48.2%; Pred. No. 3.4e-05;
Matches 279; Conservative 0; Mismatches 289; Indels 11; Gaps 2;

QY 523 TTTTATTCGTTACTGCTCATATTCAATTAATTCATTAATAATAACAGTCTTAAAGGC 582
DB 15055 TATATATATATATTTTATTATTATTATTATTATTATTATTATTATTATTATTATT 15114

QY 583 TGTTTATGGATGAATATTCGAAATTAATCATCAATAATTCATGCTATTATTACTTGCTG 642
DB 15115 TAAATTAATTAATTAATATATATATTTAAATTAATTAATTAATTAATTAATTAATG 15174

QY 643 TATTGGTATCAACTTCATGCTCTATACATGTAATAATTTTCGAGTTAGACCTTAATCA 702
DB 15175 TATTTAATTCATTTTAAATTAATTA-----ATATATATATTATTATTATTCAATTG 15226

QY 703 AGTAAATTTGCTATTTAATTTATTATCTGAATAATATGTAATCGATTGCTGTTAT 762
DB 15227 TTTTAAATTAATTAATAATAATAATATGTTTATTTTAAATTAATTTATTATAT 15286

QY 763 TTTTATGTTGTTTCATTTTAAACGCGTGAGCTTGCAATTCATATTTTATGATGA 822
DB 15287 ATTATTATATATATATATATTTTATTTTAAATTAATTTATTATTATGTTATA 15346

QY 823 CAACATCTTTGATGAAGTATTTAAGATATTGTTTAATGATGAGGGTTTGGTCTATTTT 882
DB 15347 TTTTATTTTATTTCCATTTTACATTTTATATATATATATTTTATAAATGTTAACT 15406

QY 883 TTATATTAATCAATAATAAATCAACAATATAT---GTATTTTGTGCTTTTTATAGTG 939
DB 15407 TTATATAAAAATTTATATATATATTAATAATATATTATTAGTAGTAATAATTTATTAT 15466

QY 940 TTTCTTTTAAAGAGGTAGGATGACCTTAAGGTCGCTAAATATATGCGTAATTTGCCATTGC 999
DB 15467 CATCTTTAAATAATTCATTTTAACTAAATGATTATTAAATATTTTATATAGAGATAAGA 15526

QY 1000 TATAATTCACCTCAAGATACACTATTGGCAAAATGACAAATATGTCACATTCGATGAAA 1059
DB 15527 TATGATTCAAATGTAAATAATATATTATTGAACTTTAAAGATTTTATAATATGTTTAAA 15586

QY 1060 CAATATTAGTAGATGTTGTTTTTTCGTCGCAAAATAAAAA 1098
DB 15587 CAAAAATTTGACATAATAATCTTCTTATTACAGTAATAA 15625

```







**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 2, 2004, 14:05:42 ; Search time 1304 Seconds  
(without alignments)  
11689.060 Million cell updates/sec

Title: US-10-780-347-1

Perfect score: 3588

Sequence:

1 gtcgacttattgcatgatg.....gatnaaatgagcgcgttaa 3588

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*

- 1: Geneseqn1980s:\*
- 2: Geneseqn1990s:\*
- 3: Geneseqn2000s:\*
- 4: Geneseqn2001as:\*
- 5: Geneseqn2001bs:\*
- 6: Geneseqn2002s:\*
- 7: Geneseqn2003as:\*
- 8: Geneseqn2003bs:\*
- 9: Geneseqn2003cs:\*
- 10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	72.8	2.0	11422	6	ABK39937 Human che
2	72.8	2.0	11422	6	ABL32219 Human inn
3	72.4	2.0	8093	7	ABZ10178 Haematopo
4	72.4	2.0	8093	9	ADBS4268 Pretreat
5	72.4	2.0	8093	9	ADBS4190 Human lym
6	71.4	2.0	61020	4	AAS46788 Tumour su
7	69.8	1.9	8666	9	ADBS4239 Pretreat
8	69.8	1.9	8666	9	ADE84177 Human lym
9	69.4	1.9	8093	4	AAS46436 Tumour su
10	69.4	1.9	8093	6	ABK33974 Human DNA
11	69.4	1.9	8093	6	ABL92237 Chemical
12	69.4	1.9	8093	6	ABL49332 Human pol
13	69.4	1.9	8093	7	ABZ10032 Haematopo
14	69.4	1.9	8093	7	ADA20361 Prostate
15	69.4	1.9	8093	7	ADA84168 Human ren
16	69.4	1.9	8093	9	ADBS4140 Pretreat
17	69.4	1.9	8093	9	ADE84114 Human lym
18	69	1.9	5504	6	ABL70350 Chemical
19	69	1.9	5504	6	AAS61309 Human gen
20	68.2	1.9	83391	6	ABQ57094 Human ang
21	68	1.9	6306	4	AAS45516 Chemical
22	68	1.9	6306	4	ABK28458 DNA trans
23	68	1.9	13131	6	ABL92249 Chemical

C	24	66.8	1.9	8056	7	ABZ10246	Abz10246 Haematopo
	25	66.6	1.9	8666	4	AAS46305	Aas46305 Tumour su
	26	66.6	1.9	8666	6	ABL32396	Ab132396 Human inn
	27	66.6	1.9	8666	6	ABK34008	Abk34008 Human DNA
	28	66.6	1.9	8666	6	ABQ67177	Abq67177 Human ang
	29	66.6	1.9	8666	9	ADBS4111	Adbs4111 Pretreat
	30	66.6	1.9	8666	9	ADBS4101	Adbs4101 Human lym
	31	66.6	1.9	9504	6	ABK28407	Abk28407 DNA trans
	32	66.2	1.8	17848	4	AAS45323	Aas45323 Chemical
	33	66.2	1.8	17848	6	ABK39976	Abk39976 Human che
	34	66.2	1.8	17848	6	ABK28164	Abk28164 DNA trans
	35	66.2	1.8	18624	6	ABL33702	Ab133702 Human inn
	36	66	1.8	516	7	ABX40620	Abx40620 Bovine BS
C	37	66	1.8	19124	2	AT72882	Aa72882 Plasmodiu
	38	66	1.8	19124	3	AAT98287	Aat98287 Plasmodiu
C	39	65.8	1.8	7849	6	ABL92279	Ab192279 Chemical
	40	65.8	1.8	7849	6	ADL22330	Adl22330 Chemical
	41	65.8	1.8	19459	6	ABK31212	Abk31212 Signal tr
	42	65.8	1.8	19459	6	ABL70527	Ab170527 Chemical
	43	65.4	1.8	1179	6	ABQ49604	Abq49604 Oligonuc1
C	44	65.4	1.8	1179	6	ABQ49605	Abq49605 Oligonuc1
	45	65.2	1.8	6040	6	ABL92300	Ab192300 Chemical

#### ALIGNMENTS

##### RESULT 1

ABK39937  
ID ABK39937 standard; DNA; 11422 BP.

XX

AC ABK39937;

XX

DT 21-MAY-2002 (first entry)

XX

DE Human chemically pretreated gene sequence #9 strand 2.

XX

KW Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;  
cytostatic; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;  
UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.

OS Homo sapiens.

XX

PN WO200202806-A2.

XX

PD 10-JAN-2002.

XX

PF 29-JUN-2001; 2001WO-EP007470.

XX

PR 30-JUN-2000; 2000DE-01032529.

XX

PR 01-SEP-2000; 2000DE-01043826.

XX

(EPIG-) EPIGENOMICS AG.

XX

Olek A, Piepenbrock C, Berlin K;

XX

DR WPI; 2002-154757/20.

XX

New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,  
useful for detecting cytosine methylation state of genes associated with  
pharmacogenomics and for therapy of diseases e.g. cancer.

PS Claim 1; SEQ ID NO 18; 24pp; English.

XX

The invention relates to a nucleic acid comprising a sequence at least 18  
bases in length of a segment of the chemically pretreated DNA of genes  
associated with pharmacogenomics according to one of the sequences of the  
genes ALDH6 (NM\_000693), CYP11A (NM\_000781), CYP11B1 (NM\_000497), CYP3A3  
(NM\_000776 and NM\_017460), DPYD (NM\_001110), EPHX2 (NM\_001979), OCLN  
(NM\_002538), TXNRD1 (NM\_003360), UGT8 (NM\_003360), MRP (NM\_004996),  
CC NM\_019900, NM\_019901, NM\_019902, NM\_019962, NM\_019898, NM\_019899, and  
their complementary sequences, or a sequence (S1) chosen from 87  
sequences and their complements. The chemical pretreatment is bisulphite

CC treatment to convert cytosines (but not methyl-cytosines) into uracils.  
 CC Also included are an oligomer (II) in particular an oligonucleotide or a  
 CC peptide nucleic acid (PNA)-oligomer, comprising in each case at least one  
 CC base sequence having a length of 9 nucleotides which hybridises to or is  
 CC identical to a chemically pretreated DNA of genes associated with  
 CC pharmacogenomics and their complements, arranged in an array for  
 CC analysing diseases associated with the methylation state (CpG) and/or  
 CC detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The  
 CC oligomers may also be used as PCR primers. The set of 87 nucleic acids  
 CC and their complements is useful for diagnosis and therapy of solid  
 CC tumours and cancer. The present sequence represents one the 87 DNA  
 CC sequences or its complement. Note: the sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX Sequence 11422 BP; 3913 A; 59 C; 1856 G; 5594 T; 0 U; 0 Other;

Query Match 2.0%; Score 72.8; DB 6; Length 11422;  
 Best Local Similarity 47.7%; Pred. No. 1.2e-06;  
 Matches 244; Conservative 0; Mismatches 267; Indels 1; Gaps 1;  
 QY 605 AATTATCATCAATAAATTGATGCTATTATTACTTCCTGTTATGTTATCAACTTTTCATGCT 664  
 Db |||||  
 QY 665 CTATACATGTAATATATTCGAGTTAGACCTTAATTCGAGTAATTTGCTATTATTA 724  
 Db |||||  
 QY 725 TTATCTGAATAATATGTAATCGAATTCGCTGTTATTTTATGTTTTCATTTT 784  
 Db |||||  
 QY 785 ATGACGGTGCAGCTTCGCAATTCATATTTTATGATGACCAATCTTCTATTTAATTA 844  
 Db |||||  
 QY 845 AAGATATTTGTAATCGAGGGTTTCGCTGTTATTTTATGATGACCAATCTTCTATTTAATTA 903  
 Db |||||  
 QY 904 TCAACAAATATATGTTATTTGCTTTTATAGTCTTTTAAAGAGGTTAGGATGACC 963  
 Db |||||  
 QY 964 TAAAGGTCGCTAAATATATGCGTAAATTCCTATTAATTCCTCAAGATACACT 1023  
 Db |||||  
 QY 1024 ATTGCAAAATGACAAATATGTCATCTCGTATGAACAATATTAGTAGATGTTGTTT 1083  
 Db |||||  
 QY 1084 CTGCAAAATATAAATTTTCTGGTTGAAATA 1115  
 Db |||||  
 QY 10140 TTGGAATAATTAATACGATGGAATAAGTA 10171  
 Db |||||

RESULT 2  
 ABL32219  
 ID ABL32219 standard; DNA; 11422 BP.  
 AC ABL32219;  
 XX  
 XX 26-MAR-2002 (first entry)  
 DT  
 XX Human immune system associated gene SEQ ID NO: 192.  
 DE  
 XX Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;

KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;  
 KW ds.  
 XX Homo sapiens.  
 XX WO200200928-A2.  
 XX 03-JAN-2002.  
 XX 02-JUL-2001; 2001WO-EP007537.  
 XX 30-JUN-2000; 2000DE-01032529.  
 XX 01-SEP-2000; 2000DE-01043826.  
 XX (EPIG-) EPIGENOMICS AG.  
 XX Olek A, Piepenbrock C, Berlin K;  
 XX WPI; 2002-130909/17.  
 XX Nucleic acid comprising fragment of chemically modified gene, useful for  
 XX diagnosis and treatment of diseases associated with abnormal cytosine  
 XX methylation.  
 XX Claim 1; SEQ ID NO 192; 32pp + Sequence Listing; German.  
 XX The present invention provides a number of human immune system associated  
 XX genes which are modified by the methylation of cytosines. The sequences  
 XX can be used in the diagnosis and treatment of immune system disorders,  
 XX including eye diseases such as retinopathy, neovascular glaucoma and  
 XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 XX diseases. The present sequence is a gene of the invention  
 XX  
 XX Sequence 11422 BP; 3913 A; 59 C; 1856 G; 5594 T; 0 U; 0 Other;

Query Match 2.0%; Score 72.8; DB 6; Length 11422;  
 Best Local Similarity 47.7%; Pred. No. 1.2e-06;  
 Matches 244; Conservative 0; Mismatches 267; Indels 1; Gaps 1;  
 QY 605 AATTATCATCAATAAATTGATGCTATTATTACTTCCTGTTATGTTATCAACTTTTCATGCT 664  
 Db |||||  
 QY 665 CTATACATGTAATATATTCGAGTTAGACCTTAATTCGAGTAATTTGCTATTATTAATTA 724  
 Db |||||  
 QY 725 TTATCTGAATAATATGTAATCGAATTCGCTGTTATTTTATGATGACCAATCTTCTATTTAATTA 784  
 Db |||||  
 QY 785 ATGACGGTGCAGCTTCGCAATTCATATTTTATGATGACCAATCTTCTATTTAATTA 844  
 Db |||||  
 QY 845 AAGATATTTGTAATCGAGGGTTTCGCTGTTATTTTATGATGACCAATCTTCTATTTAATTA 903  
 Db |||||  
 QY 904 TCAACAAATATATGTTATTTGCTTTTATAGTCTTTTAAAGAGGTTAGGATGACC 963  
 Db |||||  
 QY 964 TAAAGGTCGCTAAATATATGCGTAAATTCCTATTAATTCCTCAAGATACACT 1023  
 Db |||||  
 QY 1024 ATTGCAAAATGACAAATATGTCATCTCGTATGAACAATATTAGTAGATGTTGTTT 1083  
 Db |||||





Db 6334 AAAATGATATAGTATTAGATTGTTAGATGTTAGAAATGCAATTAATTTAAATTTGGA 6393  
QY 636 CTTCGTGTAATGGTATCAACTTTCATGCTCTATACATGTAATATATTTCAGTTAGACCT 695  
Db 6394 ATTGTTGATATATTTTATATGTAAGATAGTATATAGTAATATTTAAAGTAGTTT 6453  
QY 696 TAATTCAGGTAATTTGCTATTTAATATATCTGAATATATGTAATCGATGCTGTTG 755  
Db 6454 TAATTATAGATTGTAAGTAATTTGTAATTTTATTAAGATAATTTGTTTGTATAAATA 6513  
QY 756 TGGTATTTTATGTTTGTTCATTTTAAATGACGGTGACCTCTGCAATCATATTTT 815  
Db 6514 GTAATTTTAAATTTTGTATTTATGAAAGGTAATTTTAAAGTTTATTTATGTAAT 6573  
QY 816 ATGATGACAACATCTTTCATGTAAGTATTTAAGATATTTTAAATGATGAGGGTTGCGT 875  
Db 6574 AATTATAAATAGGATTTAAATTTATTTATATAGATTTTAAATGATTTATATAATTTAA 6631  
QY 876 GTATTTTATATTAATCATATAAATCAACAATATATTTTGTGTTGCTTTTAT 935  
Db 6632 AAATTTTGTGTTTAAAGTTTGTGTTTAAATTTGAAATATTTTAAATTTGTTGTTAAT 6691  
QY 936 AGTGTCTCTTTAAAGAGGTAGGATGACCTAAAGTGCCTAAATATGCGGTAATTTGCCA 995  
Db 6692 TTACTAAATTTAAGGAATTAGTTATGATTAG-----AATTAGTGGTAATGATG 8743  
QY 996 TTGCTATATTCACCTCAAGATACACTATGCGAATTTGCAAAATGCAAAATGATCCTCGTAT 1055  
Db 6744 TTTGGGTAATTAATAAATGATGTTAAGATTTTAAAGATTTTAAAGATTTGTTAGTGT 6803  
QY 1056 GAAACATATATAGTAGATGTTGTTTTCGTCGCAAAATTAATTTTCTGTT 1109  
Db 6804 TAATTAGAGTAGTAATTTGTTTATTTAGAGGATTTGAGAAATGTTAGGGGT 6857

RESULT 6  
AAS46788  
ID AAS46788 standard; DNA; 61020 BP.  
XX  
AC AAS46788;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Tumour suppressor gene derived chemically modified sequence #514.  
XX  
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;  
KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;  
KW cytosine methylation; ds.  
XX  
OS Homo sapiens.  
XX  
XX WO200168912-A2.  
XX  
XX 20-SEP-2001.  
XX  
XX 15-MAR-2001; 2001WO-EP002955.  
XX  
XX 15-MAR-2000; 2000DE-01013847.  
XX  
XX 06-APR-2000; 2000DE-01019058.  
XX  
XX 07-APR-2000; 2000DE-01019173.  
XX  
XX 30-JUN-2000; 2000DE-01032529.  
XX  
XX 01-SEP-2000; 2000DE-01043826.  
XX  
XX (EPIG-) EPIGENOMICS AG.  
XX  
XX Olek A, Piepenbrock C, Berlin K;  
XX  
XX WPI; 2001-602752/69.  
XX  
XX Fragments of chemically modified genes associated with tumor suppressor  
XX PT genes and oncogenes, useful in designing primers and probes for analyzing  
XX PT diseases associated with cytosine methylation state e.g. cancer.  
XX

PS Claim 1; SEQ ID NO 514; 27bp; English.  
XX  
CC The invention relates to a nucleic acid comprising a sequence of 18  
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with  
CC bisulphite, of genes associated with tumour suppression and oncogenes  
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and  
CC 500 are missing from the sequence listing) sequences (Ss) and sequences  
CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-  
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of  
CC probes for detecting the cytosine methylation state and/or single  
CC nucleotide polymorphisms and also to be used in an array for analysing  
CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The  
CC probes can also be used in a method for ascertaining genetic and/or  
CC epigenetic parameters for the diagnosis and/or therapy of existing  
CC diseases or the predisposition to specific diseases, by analysing  
CC cytosine methylations. The parameters may be compared to another set of  
CC genetic and/or epigenetic parameters, the differences serving as basis  
CC for diagnosis and/or prognosis events, which are disadvantageous to  
CC patients. The present sequence is one of the 533 genomic sequences  
CC derived from tumour suppressor genes and oncogenes. Sequences with even  
CC numbered Seq ID numbers are the complementary sequence of the  
CC corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID  
CC 535, except for those whose partner sequence is missing). Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 61020 BP; 17884 A; 634 C; 12537 G; 29965 T; 0 U; 0 Other;  
Query Match 2.0%; Score 71.4; DB 4; Length 61020;  
Best Local Similarity 48.2%; Pred No. 4.6e-06;  
Matches 201; Conservative 0; Mismatches 216; Indels 0; Gaps 0;  
QY 517 TGGTCTTAATTCGTTACTGCTCATATTCAAATTAATTAATTAATTAATTAATTAATTA 576  
Db 11783 TGTATTATATAATTTGTTATTTTATATTTATTTTATTTTATTTTATTTTATTTT 11842  
QY 577 AAAGGCTGTTTATTCGATGAATATTCGAAATATACACATAATTAATGATGCTATTATAC 636  
Db 11843 GGAAATTAATAGGTTTTTAAAGTTTAAATATTTTAAATTAATTAATTAATTTT 11902  
QY 637 TTGCTGTAATGGTATCAACTTTTCATGCTCTATACATGTAATATTTTCGAGTTAGACCTT 696  
Db 11903 TAAATTTTGTGTTTTTGTATTTTATTTTAAAGTAATATATAATTTTAAATTAAT 11962  
QY 697 AATTCAGGTAATTTGCTATTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 756  
Db 11963 GTTGTGAGAAAGTTTGGGATTTTTTTTTTTTTTTTTTTTTTTTATTTTATTTGTTAT 12022  
QY 757 GGTATTTTTTATGTTTGTTCATTTTTTAATCAGCGTGAGCTTGTGCATTCATATTTT 816  
Db 12023 TGTATAATTTATTTTATTTTATTTTGTATTTTATTTTGTATTTATTTATTTT 12082  
QY 817 TGATGACAAATCTTTTGATGAAGTATTTAAGATATTTGTAATGATGAGGGTTGCGTG 876  
Db 12083 TGTAGTTTATTTTTTTTTTTTTTTTTTTTTTTTAAATTTTAGAGTAATTTTATTTT 12142  
QY 877 TATTTTTTATATAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATTT 933  
Db 12143 TGTATTATATATTTTTTTTTTTTTTTTTTTTAAAGTTTATTTTTCGTTTGAATTTGTTAT 12199  
RESULT 7  
ADB54239  
ID ADB54239 standard; DNA; 8666 BP.  
XX  
AC ADB54239;  
XX  
DT 04-DEC-2003 (first entry)  
XX  
DE Pretreated genomic DNA region 163.  
XX  
KW colon cell proliferative disorder; non methylated CpG dinucleotide;

KW	cytosatic; cancer; adenoma; carcinoma; cytosine methylation state; ds.
XX	Unidentified.
OS	WO2003072821-A2.
XX	04-SEP-2003.
PN	27-FEB-2003; 2003WO-EP002035.
XX	27-FEB-2002; 2002EP-00004551.
FD	(EPIG-) EPIGENOMICS AG.
XX	Adorjan P, Burger M, Maier S, Nimrich I, Becker E, Lesche R;
XX	Rujan T, Schmitt A;
XX	WPI; 2003-731620/69.
XX	Detecting and differentiating between colon cell proliferative disorders
XX	associated with a gene or its regulatory regions comprises contacting a
XX	target nucleic acid in a biological sample obtained from the subject with
XX	a reagent.
XX	Claim 32; SEQ ID NO 295; 74pp; English.
XX	The invention relates to a novel method for detecting and differentiating
XX	between colon cell proliferative disorders associated with at least one
XX	gene or its regulatory regions. The method comprises contacting a target
XX	nucleic acid in a biological sample obtained from the subject with at
XX	least one reagent or a series of reagents, where the reagent or series of
XX	reagents, distinguishes between methylated and non methylated CpG
XX	dinucleotides within the target nucleic acid. The molecules of the
XX	invention demonstrate cytostatic activity whilst the method may useful
XX	for detecting and differentiating between colon cell proliferative
XX	disorders, including cancers such as colon adenoma and colon carcinoma.
XX	The PNA (peptide-nucleic acid)-oligomers are useful as probes for
XX	determining cytosine methylation state or single nucleotide
XX	polymorphisms. The current sequence is that of the pretreated genomic DNA
XX	region of the invention. This sequence is not shown within the
XX	specification but is taken from Wipoweb.
XX	Sequence 8666 BP; 2444 A; 0 C; 1868 G; 4354 T; 0 U; 0 Other;
SQ	Query Match 1.9%; Score 69.8; DB 9; Length 8666;
	Best Local Similarity 45.0%; Pred. No. 5.7e-06;
	Matches 263; Conservative 0; Mismatches 322; Indels 0; Gaps 0;
QY	525 TAATTCGTTAGTGTGCTATTCATTAATTAATTCACATTAATAAATACAGTCTTAAAGGCTG 584
DB	30 TAATTAGTGTTTATTGTTGTTGTTGTTTATTATTATTATTAGTGGTTTATTGTTTGGTT 89
QY	585 TTTATTGGATGAATATTCGAAATATATACATAAATTAATGATGCTATTACTTCTGTTA 644
DB	90 TTAAGAATAAAATAAAGGTTTTTTTGTGAGTTTATTATTATTATTATTATTATT 149
QY	645 TTGGTATCACTTCATCTCTATACATGTAATATATTCAGTTAGCTTAATCAAG 704
DB	150 ATAAGAGGATTTTAAATTTTATGATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 209
QY	705 GTAAATTTGCTATTTTAAATTTATCTGAATATATGTAATGATGCTTTTGCTTTATT 764
DB	210 TTTTGTAAATGATTTTAAATTTTATTATTATAGAGGAAGTGTTTTAAATTTTATTGTT 269
QY	765 TTAATGTTTGTTCATTTTAAAGCGGTGACGTGTGCAATTCATATTTTATGATGACA 824
DB	270 TTGATTTTGTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 329
QY	825 ACATCTTTGATGATTTTAAAGATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 884
DB	330 GTATTTTGAGTTTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 389
QY	885 ATATTAAATCATATAAATAATCAATATATGTTATTATTGTTCTTTATTAGTGTCTT 944

DB	390 ATTGTTTGTGTTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 449
QY	945 TTAAGAGGTAGGATGACCAAGTGCCTAATATGGCTAATATGGCAATTCCTATAA 1004
DB	450 GGTATTTGTTTATTTAGGTATATGTTGTTTATTTTATTTTATTTTATTTTATTTAT 509
QY	1005 TTCACCTCAAGATACACTATTTGCCAAATGACAAATATGCACTTCGTTATGAAACAATA 1064
DB	510 TTTATGGTATAGGTTTGTGATTTATTAATTAAGTTATGAAGAGTAGTATGTTAGTG 569
QY	1085 TTAGTAGATGTTGTTTGTGCTGCAAAATATAAATTTTCTGGTT 1109
DB	570 TTTTATTTTGTGTTATTTTATTTTATTTTATTTTATTTTGTGTT 614
RESULT 8	
ADE84177	
ID	ADE84177 standard; DNA; 8666 BP.
XX	AC ADE84177;
XX	29-JAN-2004 (first entry)
DE	Human lymphoid cell proliferative disorder gene derived DNA #113.
KW	ds; lymphoid cell proliferative disorder; methylation;
KW	methylated CpG dinucleotide; single nucleotide polymorphism; SNP;
KW	diffuse large B-cell lymphoma; mantle cell lymphoma;
KW	chronic lymphocytic leukemia; small lymphocytic lymphoma;
XX	follicular lymphoma; diagnosis; prognosis.
OS	Homo sapiens.
XX	WO2003044226-A2.
XX	30-MAY-2003.
XX	25-NOV-2002; 2002WO-EP013265.
XX	23-NOV-2001; 2001DE-01057491.
XX	28-DEC-2001; 2001DE-01064501.
XX	(EPIG-) EPIGENOMICS AG.
PA	Burger M, Caldwell C, Genc B, Becker E, Maier S, Nimrich I;
PI	WPI; 2003-457621/43.
DR	Detecting and differentiating between lymphoid cell proliferative
XX	disorders comprises contacting a target nucleic acid with at least one
PT	reagent that distinguishes between methylated and non-methylated CpG
PT	dinucleotides.
XX	Claim 26; SEQ ID NO 173; 448pp; English.
PS	The invention relates to a method of detecting and differentiating
XX	between lymphoid cell proliferative disorders associated with at least
CC	one gene and/or their regulatory regions in a subject by contacting a
CC	target nucleic acid in a biological sample obtained from the subject with
CC	at least one reagent or series of reagents that distinguish between
CC	methylated and non-methylated CpG dinucleotides within the target nucleic
CC	acid. The genes and/or their regulatory regions are preferably selected
CC	from MDR1, CSNK2B, EGR4, AR, CDK4, RBL2, CDC35A, GPIIb beta, MYO1A, CDH3,
CC	MYO1A, ELK1, ABL1, APC, BCL2, CDH1, CDKN1A, CDKN2A, CDKN2B, FOS,
CC	GSTP1, HIC-1, MGMT, MLH1, MMS, MYC, PTEN, RBL2, TGFBR2, TP73, CDKN1C,
CC	CDK3beta, BSR1, ARAFI, BAK1, BAX or HOXA5, oligomers, peptide nucleic
CC	acid (PNA)-oligomers and/or isolated nucleic acids based on the sequences
CC	of the genes are useful for detecting the methylation state of all the
CC	CpG dinucleotides within one or more the sequences, or their complements,
CC	for determining the cytosine methylation state and/or single nucleotide
CC	polymorphisms (SNPs), and for differentiating at least two of the medical
CC	conditions such as diffuse large B-cell lymphoma, mantle cell lymphoma,



CC chronic lymphocytic leukemia, small lymphocytic lymphoma and follicular  
CC lymphoma. They are also useful for detecting of a predisposition to,  
CC differentiation between subclasses, diagnosis, prognosis, treating and/or  
CC monitoring of lymphoid cell proliferative disorder. This sequence  
CC represents a nucleic acid of a pretreated genomic DNA derived from the  
CC above mentioned genes.  
XX  
SQ Sequence 8666 BP; 2444 A; 0 C; 1868 G; 4354 T; 0 U; 0 Other;

Query Match 1.9%; Score 69.8; DB 9; Length 8666;

Best Local Similarity 45.0%; Pred. No. 5.7e-06;

Matches 263; Conservative 0; Mismatches 322; Indels 0; Gaps 0;

```
QY 525 TAATTCGTAGCTCATATTCATTAATTCACATTAATTAATAACACATTCCTAAAGCGTG 584
    |||||
Db 30 TAATTAGTGTTTATTTGTTTATGATTTTATTTTAGTGGTTTATTTAGTTGGTT 89
QY 585 TTTATTGGATGAATATTCGAATATACACATAAATGATGCTATTATTACTTGCCTGA 644
    |||||
Db 90 TTAAGAATAAATAAAGGTTTTTTTGTGTTGAGTTTTATTATTTATTTTATTT 149
QY 645 TTGGTATCACTTCATGCTCTACATGATGTAATATATTCGAGTAGACCTTAATCAAG 704
    |||||
Db 150 ATAAGAGGATATTTAAATTTTATGATTTTAAATTTTAAATGATTTATTTATTTA 209
QY 705 GTAATTGCTATTAATTAATTAATCTGAATAATGTAATCGATGCTTTGTGGTTATTT 764
    |||||
Db 210 TTTTGTAAAGTTAATTTTATTTATAGAGGAAGTGTTTAAATTTTGTGTT 269
QY 765 TTATGTTGTTTCAATTTTAAATGAGGTTGCTCATATATTTTATGATGACA 824
    |||||
Db 270 TTGATTTTATTTTAAATTTTATGTTATTTTGTGATAAAATTAATAATTTGTT 329
QY 825 ACATCTTTGATGAGTATTAAGATATTTGTAATGCAATGAGGGTTTCGCTGATTTT 884
    |||||
Db 330 GTATTTGAGTTTATTTTATTTATTTTGTAGGGAATTTTAAATTTTAAATAATTT 389
QY 885 ATATTAATCATATAAATAACAAATATATGTTATTTTGTCTTTTATAGTGTTCTT 944
    |||||
Db 390 ATTTGTTGTTGATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 449
QY 945 TTAAGAGGTAGATGACCTAAAGGTCGCTCAATATGCGGTAAATGCGCATTCGTATAA 1004
    |||||
Db 450 GGTATTTGTTTATTTAGGTATATGTTGTTAGGTTTATTTTATTTTATTTTATTTAT 509
QY 1005 TTCACCTCAAAGATACACTATTCGCAATTCGCAATATGCTCACTTCGTATGAACAATA 1064
    |||||
Db 510 TTTATGATATAGGTTTGTGAATTAATTAATTAATTAATTAATTAATTAATTAATG 569
QY 1065 TTAGTAGATGTTGTTTGTGCTGCAAAATAAATTTTCTGTT 1109
    |||||
Db 570 TTTTATTTTGTGTTATTTTATTTATTTTAGTTTATTTTGTGTT 614
    |||||
```

## RESULT 9

AA546436 standard; DNA; 8093 BP.

AA546436;

18-DEC-2001 (first entry)

Tumour suppressor gene derived chemically modified sequence #158.

Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;  
tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;  
cytosine methylation; ds.

Homo sapiens.

W0200169912-A2.

20-SBP-2001.

15-MAR-2001; 2001WO-BP002955.

15-MAR-2000; 2000DE-01013847.

06-APR-2000; 2000DE-01019058.

07-APR-2000; 2000DE-01019173.

30-JUN-2000; 2000DE-01032529.

01-SEP-2000; 2000DE-01043826.

(BPIG-) EPIGENOMICS AG.

Olek A, Piepenbrock C, Berlin K;

WPI; 2001-602752/68.

Fragments of chemically modified genes associated with tumor suppressor  
genes and oncogenes, useful in designing primers and probes for analyzing  
diseases associated with cytosine methylation state e.g. cancer.

Claim 1; SEQ ID NO 158; 27pp; English.

The invention relates to a nucleic acid comprising a sequence of 18  
bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with  
bisulphite, of genes associated with tumour suppression and oncogenes  
having a sequence taken from 536 (actually 533 since numbers 406, 458 and  
500 are missing from the sequence listing) sequences (SS) and sequences  
complementary to (SS). The nucleic acid may be a peptide nucleic acid-  
oligomer (PNA) of at least 9 nucleotides and may form part of a set of  
probes for detecting the cytosine methylation state and/or single  
nucleotide polymorphisms and also to be used in an array for analysing  
diseases associated with CpG dinucleotides e.g. cancers and tumours. The  
probes can also be used in a method for ascertaining genetic and/or  
epigenetic parameters for the diagnosis and/or therapy of existing  
diseases or the predisposition to specific diseases, by analysing  
cytosine methylations. The parameters may be compared to another set of  
genetic and/or epigenetic parameters, the differences serving as basis  
for diagnosis and/or prognosis events which are disadvantageous to  
patients. The present sequence is one of the 533 genomic sequences  
derived from tumour suppressor genes and oncogenes. Sequences with even  
numbered seq ID numbers are the complementary sequence of the  
corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID  
535, except for those whose partner sequence is missing). Note: The  
sequence data for this patent did not form part of the printed  
specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pct\_sequences

Sequence 8093 BP; 2419 A; 175 C; 1689 G; 3810 T; 0 U; 0 Other;

Query Match 1.9%; Score 69.4; DB 4; Length 8093;

Best Local Similarity 47.5%; Pred. No. 7e-06;

Matches 277; Conservative 0; Mismatches 296; Indels 10; Gaps 2;

QY 527 ATTCGTTACTGCTCATATTCATTAATTCACATATTAATAACAGTCTTAAAGCGTCTT 586  
 |||||  
Db 6285 ATTTGTTGTTTATTTTAAATAATTTGTTTAAAGTAAATATTTTGTTTAAATGATATA 6344  
 |||||

QY 587 TATTCGATGAATATTCGAAATATTCACATAATATTCATGCTATTTATTTACTTGTGTTAT 646  
 |||||  
Db 6345 GTATTAGATTTTGTAGATGTTAGAAATGGATTTATTTTAAATTTGGAATTCGTATA 6404  
 |||||

QY 647 GGTATCAACTTTCATGCTCTATACATGTAATATTTTCGAGTTAGACCTTAATTCAGGT 706  
 |||||  
Db 6405 TATTTATATGTAAGATAGTATATAAGTAGAAATATTTTAAAGTAGTATTTATTTATAGAT 6454  
 |||||

QY 707 AATTGTTCTATTTATTTATTTATTCGAAATATTCGATATTCGATTCGTTTGTGGTTATTTT 766  
 |||||  
Db 6465 TGTAGTAAATTTGTTATTTTATTTAAGATATTTGTTTGTGTTAAATAGTAAATTTTAA 6524  
 |||||

QY 767 ATGTTGTTTCATTTTAAAGACGGTGGCTTGTGCATTCATATTTTATGATGACAAC 826  
 |||||  
Db 6525 ATTTTGTGTTTATGAAAGGTAATTTTAAAGTTTATTTAAAGTTTATTTGTAATAATTAATA 6584  
 |||||

QY 827 ATCTTTGATGAAGTATTTAGATATTTGTAATGATGAGGGGTTTGGGTGTTATTTTAT 886  
 |||||



ID ABL92237 standard; DNA; 8093 BP.  
 AC ABL92237;  
 XX  
 DT 01-JUL-2002 (first entry)  
 XX  
 DE Chemically treated DNA repair gene fragment complementary to#23.  
 XX  
 KW DNA repair; cytosine methylation; PMS2L1; PMS2L2; PMS2L3; PMS2;  
 KW L4; PMS2L5; PMS2L6; MGMT; MSH2; NUDT1; TDG; INPPL1; RFC4; DDI1L; FANCB;  
 KW XRC8; ataxia telangiectasia; aging; Bloom's syndrome; Cockayne syndrome;  
 KW Nijmegen breakage syndrome; Werner syndrome; immunodeficiency;  
 KW trichiodystrophy; Fanconi's anaemia; solid tumour; cancer; ds.  
 XX  
 OS Unidentified.  
 XX  
 XX WO200181622-A2.  
 PN  
 XX  
 PD 01-NOV-2001.  
 XX  
 XX 06-APR-2001; 2001WO-EP003972.  
 XX  
 XX 06-APR-2000; 2000DE-01019058.  
 PR  
 PR 07-APR-2000; 2000DE-01019173.  
 PR  
 PR 30-JUN-2000; 2000DE-01032529.  
 PR  
 PR 01-SEP-2000; 2000DE-01043826.  
 XX  
 XX (EPIG-) EPIGENOMICS AG.  
 PA  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 XX WPI; 2002-034446/04.  
 DR  
 XX

New nucleic acid derived from genes associated with DNA repair, useful  
 for diagnosis, e.g. of ataxia telangiectasia, by determination of  
 cytosine methylation.  
 PS Claim 1; SEQ ID NO 46; 25pp + Sequence Listing; English.

The invention relates to nucleic acids containing a sequence of at least  
 18 nucleotides of chemically treated DNA of genes associated with DNA  
 repair, and their complements. The invention also relates to nucleic  
 acids comprising at least 18 base pairs of the chemically pretreated DNA  
 of genes associated with DNA repair selected from PMS2L1, PMS2L2,  
 PMS2L3, PMS2, L4, PMS2L5, PMS2L6, MGMT, MSH2, NUDT1, TDG, INPPL1,  
 RFC4, DDI1L, FANCB, or XRC8. Nucleic acids of the invention and related  
 oligomers, are useful for diagnosis of diseases associated with gene  
 repair, specifically ataxia telangiectasia, aging, Bloom's syndrome,  
 Cockayne syndrome, Nijmegen breakage syndrome or Werner syndrome,  
 immunodeficiency, trichiodystrophy, Fanconi's anaemia, solid tumours  
 and cancer, particularly by determining status of cytosine methylation  
 and/or by detecting single-nucleotide polymorphisms. Determination of  
 individual methylation patterns may allow development of individualised  
 therapies. The sequences given in records ABL92192-ABL92335 represent  
 chemically pre-treated DNA fragments from genes associated with DNA  
 repair, and their complements. Note: The sequence data for this patent is  
 not represented in the specification, but is based on sequence  
 information supplied by the European Patent Office

SQ Sequence 8093 BP; 2419 A; 175 C; 1689 G; 3810 T; 0 U; 0 Other;

Query Match 1.9%; Score 69.4; DB 6; Length 8093;  
 Best Local Similarity 47.5%; Pred. No. 7e-06;  
 Matches 277; Conservative 0; Mismatches 296; Indels 10; Gaps 2;

QY 527 ATTCGTACTGCTCATTAATTAATTCATTAATAAATACAGTCTTAAAGGCTGTT 586  
 Db 6285 ATTTGTTTATTTTAAATTTTGTGTTTAAAGTAAATATTTGTTTAAATGATATA 6344  
 QY 587 TATTGGATCAATTCGAATATATCATATAAATGATGCTATTATTACTTCTGTTATT 646  
 Db 6345 GTATTAGATTGTTAGATGTTAGAAATGGAATTTATTTTAAATTTGGAATTTGCTGATA 6404

QY 647 GGTATCAACTTTTCATGCTCTATACATGTAATATATTCAGTTAGACCTTAATCAAGGT 706  
 Db 6405 TATTTTATATGTAAGATAGTATATAGTAGAAATATTTAAAGTAGTATTTTATTATAGAT 6464  
 QY 707 AATTTTGTCTATTAAATTAATCTCGAATAATATGTAATCGATTGCTTTGTGTTATTTT 766  
 Db 6465 TGTAGTAATTTGTAATTTTATTAAGATAATTTGTTTGTGTTAAATAGTAATTTTAA 6524  
 QY 767 ATGTTTGTTCATTTTAAATGACGGTGGAGCTTGTGCATTTCATATTTTATGATGACAAC 826  
 Db 6525 ATTTTGTTTTATTAATCAAAAGGTAATTTTAAAGTTTATTAATGTAATAATTAATAATA 6584  
 QY 827 ATCTTTGATGAAGTATTTAAGATATGTTTAATGCATGAGGGCTTTCGCGTATTTTAT 886  
 Db 6585 GGATTTAATTTATTTATAGATTTTATTTTAAAGTATTAATAATTTT-AAAAATTTTGT 6642  
 QY 887 ATTAATCATATAAATAACAATATATATTTTGTCTCTTTTATGTTGTTCTTTT 946  
 Db 6643 TTAAGATTTTGTTTTAAATTTATTTGAAATATTTTAAATTTGTTTATTAATTTAGTAAAT 6702  
 QY 947 AAAGAGGTAGGATGACCTAAAGTGCCTTAATATATGCGTAAATTTGCCATTCCTATAAT 1006  
 Db 6703 AAGGAATTAGGTTATGATTAAG-----AATTTAGGTGGAATTCATGTTTGGTAAT 6754  
 QY 1007 CACCTCAAAAGATACACTATTGCAAAATTCACAAATATGCTCCTATGTAACCAATAT 1066  
 Db 6755 AAAATAAATGTTAAGAGTTTAAATAATTAAGTTGTTAGTGGTTTTTAAATTAGAGGT 6814  
 QY 1067 AGTAGATGTTGTTTGTCTGCAAAATAAATAATTTTCTGTT 1109  
 Db 6815 AGTAATTTGTTTATTTTAGAGGATGTTGAGAAATGTAGGGGT 6857

## RESULT 12

ID ABL49332  
 AC ABL49332 standard; DNA; 8093 BP.

XX  
 DT 01-MAY-2002 (first entry)  
 DE Human polynucleotide associated with DNA replication SEQ ID NO 32.  
 XX  
 KW Human; cytostatic; neuroprotective; nootropic; immunostimulant;  
 KW gene therapy; gene regulation; DNA replication; CENPB; DNA2L; ATR; CHD1L;  
 KW ERCC3; SNRPAL; RAD50; LIG2; cytosine methylation; Ataxia telangiectasia;  
 KW ATR-X; Bloom's syndrome; tumour; cancer; methylation; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200177377-A2.  
 XX  
 PD 18-OCT-2001.  
 XX  
 PF 06-APR-2001; 2001WO-EP003971.  
 XX  
 PR 06-APR-2000; 2000DE-01019058.  
 PR 07-APR-2000; 2000DE-01019173.  
 PR 30-JUN-2000; 2000DE-01032529.  
 PR 01-SEP-2000; 2000DE-01043826.  
 XX  
 XX (EPIG-) EPIGENOMICS AG.  
 XX  
 XX Olek A, Piepenbrock C, Berlin K;  
 XX  
 XX WPI; 2002-017471/02.  
 DR  
 XX  
 PT New nucleic acid sequences from chemically modified genes associated with  
 PT DNA replication, useful for analyzing cytosine methylations for diagnosis  
 XX and therapy of diseases e.g. Ataxia telangiectasia.  
 XX  
 PS Claim 1; SEQ ID NO 32; 23pp; English.





PD 27-DEC-2002.  
XX 14-JUN-2002; 2002WO-EP006603.  
XX 14-JUN-2001; 2001DE-01028509.  
XX (EPIC-) EPIGENOMICS AG.  
XX Distler J, Model F, Adorjan P;  
XX WPI; 2003-183991/18.  
XX Method for characterizing, classifying and/or differentiating renal and  
XX prostate cancers, by analyzing the genetic and/or epigenetic parameters  
XX of genomic DNA, particularly by determining its cytosine methylation  
XX status.  
XX Claim 1; Page 80-83; 21lpp; English.  
XX The invention relates to a novel method for characterising, classifying  
XX and/or differentiating renal and prostate cancer. The method comprises  
XX extracting genomic DNA from a biological sample, converting cytosine  
XX bases (by chemical treatment) that are unmethylated at the 5-position to  
XX uracil or another base, and amplifying at least one fragment of the  
XX chemically pretreated genomic DNA using sets of primer oligonucleotides  
XX and a polymerase. The method is useful for detecting the cytosine  
XX methylation state and/or single nucleotide polymorphisms in genomic DNA,  
XX particularly for characterising, classifying and/or differentiating renal  
XX and prostate cancers. The oligomers are useful as primer oligonucleotides  
XX for the amplification of any of the 112 DNA sequences of the invention.  
XX The set of oligomer probes is useful for detecting the cytosine  
XX methylation state and/or single nucleotide polymorphisms in any of the  
XX 112 chemically pretreated genomic DNA sequences. The method is also  
XX useful for identifying the tissue of origin of cancer cells. The method  
XX allows the classification, differentiation and/or diagnosis of cancer  
XX tissues using minute samples which would be inadequate for histological  
XX or cytological analysis. The present sequence represents one of the 112  
XX DNA sequences of the invention.  
XX Sequence 8093 BP; 2419 A; 175 C; 1689 G; 3810 T; 0 U; 0 Other;  
Query Match 1.9%; Score 89.4; DB 7; Length 8093;  
Best Local Similarity 47.5%; Pred. No. 7e-06;  
Matches 277; Conservative 0; Mismatches 296; Indels 10; Gaps 2;  
QY 527 ATTGCTTACTGCTCATATTCATTAATTAATCACTATTAATAAACAAGTCTTAAGCGCTGTT 586  
Db 6285 ATTGTTGTTTATTTTAAATAATTTTGTAAAGTAAATATTTGTTTAAATATGATATA 6344  
QY 587 TATTGGATGAATATTCGAATATTCACATAATAATTCATGCTATTATTACTTGTGCTATT 645  
Db 6345 GTATTAGATTTTGTAGATGTTAGAAATGGAATTTATTTTAAATTTGGAATTCGTATA 6404  
QY 647 GGTATCAACTTTCATGCTCTATACATGPAATATATTTTCGAGTTAGACCTTAATTCAGGT 706  
Db 6405 TATTTTATATGATAGATAGTATATAAGTAGAATAATTTTAAAGTAGTATTTTATTATAGAT 6464  
QY 707 AATTTGCTATTATTATTATTCGATAATAATATGTAATCGATTGCTTTGTTGTTATTTT 766  
Db 6465 TGAGTAATTTTGTATTTTATTAAGATTAATTTGTTTGTGTTAAATAGTAATTTTAA 6524  
QY 767 ATGTTTGTTCATTTTAAACCGGTGAGCTTGTGCAATCATATTTTATGATGACAAC 826  
Db 6525 ATTTTGTTTTATATGAAAGGTATTTTAAAGTTTATTATGTAATAATTAATTAATA 6584  
QY 827 ATCTTTGTAGATTTTAAAGTATTTGTAATGATGAGGGGTTTGGGTGATTTTAT 886  
Db 6585 GGATTTAATTTATATTTATAGATTTTATTAAGTATTTTATTAATTTTAAATTTTGT 6642  
QY 887 ATTAATCATATAAATAACAATATATATGTTATTTTGTCTTTTATAGTCTTTT 946  
Db 6643 TTAAGTTTGTTTTAAATTTATGAAATATTTTAAATTTGTTGTTATTAATTTAGTAATTT 6702

QY 947 AAAGAGGTAGGATGACCTAAAGGTGCGCTAAATATGGGTAATTCGATTGCTATTAATT 1006  
Db 6703 AAGGAATTAGGTTATGATTAAAG-----AATTTAGGTGGAAATGATGTTTGGGTAATT 6754  
QY 1007 CACCTCAAGATACACTATTGGCAAAATGACAAATATGTCTACTTCGTATGAAACAATATT 1066  
Db 6755 AAATAAATGGTATTAAGAGTTTAAATAATTAAGTTGTTAGTGGTTTATTAATTAGAGGT 6814  
QY 1067 AGTAGATGTTGTTTTTGTCTGCAAAAATAAAAAATTTTCTGTTT 1109  
Db 6815 AGTAATTTGTTTATTTTAGAGGATGTTGAGAAATGTTAGGGGT 6857

Search completed: July 2, 2004, 16:42:07  
Job time : 1317 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 2, 2004, 15:49:42 ; Search time 237 Seconds  
(without alignments)  
8401.535 Million cell updates/sec

Title: US-10-780-347-1  
Perfect score: 3588  
Sequence: 1 gtcgacttattgatgatg.....gatnaattgagcgcttaa 3588

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:\*

1: /cgn2\_6/prodata/2/ina/5A COMB.seq.\*  
2: /cgn2\_6/prodata/2/ina/5B COMB.seq.\*  
3: /cgn2\_6/prodata/2/ina/6A COMB.seq.\*  
4: /cgn2\_6/prodata/2/ina/6B COMB.seq.\*  
5: /cgn2\_6/prodata/2/ina/6CTUS COMB.seq.\*  
6: /cgn2\_6/prodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	69.4	1.9	8093	4	US-10-204-708-32
2	68.6	1.9	837	3	US-08-998-416-288
3	66	1.8	19124	2	US-08-487-828B-13
4	65.2	1.8	6040	4	US-10-204-708-69
5	65	1.8	658	3	US-08-998-416-595
6	63	1.8	717	4	US-09-252-991A-8972
7	63	1.8	819	4	US-09-252-991A-8861
8	60.4	1.7	5562	4	US-10-204-708-63
9	60	1.7	6866	4	US-10-204-708-20
10	59.8	1.7	640681	4	US-09-790-988-1
11	59.4	1.7	600	4	US-09-252-991A-752
12	59.4	1.7	681	4	US-09-252-991A-723
13	59.4	1.7	2019	4	US-09-252-991A-658
14	59.2	1.6	665	2	US-08-883-795A-36
15	57.6	1.6	11015	4	US-10-204-708-55
16	56	1.6	579	2	US-09-252-991A-9199
17	55.6	1.5	8920	2	US-08-446-855A-1
18	55.6	1.5	8920	3	US-09-150-741-1
19	55.2	1.5	6113	4	US-10-204-708-14
20	55	1.5	10467	4	US-10-204-708-2
21	54.6	1.5	53332	4	US-09-801-861-3
22	53	1.5	832	4	US-09-621-976-2813
23	53	1.5	6669	4	US-10-204-708-6
24	53	1.5	19124	2	US-08-487-828B-13
25	52.8	1.5	6243	2	US-09-056-075-1
26	52.6	1.5	5152	4	US-10-204-708-73
27	52.6	1.5	5501	4	US-10-204-708-38

28	52.4	1.5	7304	4	US-10-204-708-43	Sequence 43, Appl
29	52	1.4	6306	4	US-10-204-708-50	Sequence 50, Appl
30	52	1.4	6326	4	US-10-204-708-57	Sequence 57, Appl
31	51.8	1.4	8961	4	US-10-204-708-80	Sequence 80, Appl
32	51.4	1.4	396	4	US-09-640-173-53	Sequence 53, Appl
33	51.4	1.4	396	4	US-09-713-550-53	Sequence 53, Appl
34	51.4	1.4	19233	4	US-10-204-708-45	Sequence 45, Appl
35	51.2	1.4	11131	4	US-10-204-708-28	Sequence 28, Appl
36	50.8	1.4	8607	4	US-10-204-708-72	Sequence 72, Appl
37	50.6	1.4	2347	4	US-08-956-171E-153	Sequence 153, Appl
38	50	1.4	8607	4	US-10-204-708-71	Sequence 71, Appl
39	49.8	1.4	615	3	US-08-998-416-186	Sequence 186, Appl
40	49.4	1.4	354	4	US-09-328-352-2580	Sequence 2580, Appl
41	49.4	1.4	9347	4	US-10-204-708-35	Sequence 35, Appl
42	49.4	1.4	64681	4	US-09-790-988-1	Sequence 1, Appl
43	49.2	1.4	6156	4	US-10-204-708-60	Sequence 60, Appl
44	49	1.4	5666	4	US-10-204-708-29	Sequence 29, Appl
45	49	1.4	19233	4	US-10-204-708-46	Sequence 46, Appl

#### ALIGNMENTS

##### RESULT 1

US-10-204-708-32  
; Sequence 32, Application US/10204708  
; Patent No. 6677731  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication  
; TITLE OF INVENTION: by Assessing DNA Methylation  
; FILE REFERENCE: 5013.1012  
; CURRENT APPLICATION NUMBER: US/10/204,708  
; CURRENT FILING DATE: 2003-05-06  
; PRIOR APPLICATION NUMBER: PCT/EP01/03971  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: DE 10019058.8  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: DE 10019173.8  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 98  
; SEQ ID NO 32  
; LENGTH: 8093  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-204-708-32

Query Match	1.9%	Score 69.4;	DB 4;	Length 8093;
Best Local Similarity	47.5%;	Pred. No. 4.8e-08;		
Matches 277;	Conservative 0;	Mismatches 236;	Indels 10;	Gaps 2;
QY	527	ATTGCTACTGCTCATATTCATTAATTCACATAATAATTCAGTTTAAAGGCTGTT	586	
DB	6285	ATTGTTGTTTATTTTAAANATTTTGTAAAGTAATATTTGTTTAAATGATATA	6344	
QY	587	TATTGGATGAATATTCGAAATATTCACATAATAATTCATGCTATATTACTTGTGTT	646	
DB	6345	GTATTAGATTTTGTAGATGTTAGAAATGGATTATTTTAAATTTGGAATTCGATATA	6404	
QY	647	GGTATCAACTTTCATGCTCTATACATCATTAATTTTCGAGTTAGACCTTAATCAAGGT	706	
DB	6405	TATTTTATGTAAGATAGTATATTAAGTAATTTTAAAGTAGTTTTTATTATAGAT	6464	
QY	707	AATTGTCTATTTTATTTATCTCTGAATAATATGTAATCGATTCGTTTGTGTTATTTT	766	

Db 6465 TGAGTAATTTGATTTTATTATTAAGATAAATTTGTTTGTGTTAAATAGTAATTTTAA 6524  
Qy 767 ATGTTTCTTTTCATTTTAAATGACGGTGAGCTTGTGCATTCATATTTTATGATGACAAC 826  
Db 6525 ATTTTCTTTTATGTAAGAGGTAATTTTAAAGTTTATTATGTAATAATTAATTAATA 6584  
Qy 827 ATCTTTGATGAAGTATTTAGATATTTGTTTAATGATGAGGGTTTCGGTGTATTTTAT 886  
Db 6585 GGATTTAAATTTATATTTATAGATTTTAAAGTATTATATAATTTTAAATTTTGT 6642  
Qy 887 ATTAATCATAATAAATCAACAATATATGTTTATTTTGTCTTTTATAGTGTCTTTT 946  
Db 6643 TTAAGTTTGTTTTATTTATTTGAAATTTTAAATTTTGTGTTTAAATTTAGTAATTT 6702  
Qy 947 AAAGAGTAGATGACCTAAAGTTCGCTTAAATATGCGGTAATTTGCAATTCGTATAAT 1006  
Db 6703 AAGGAATTAGGTTATGATTAAG-----AATTTAGTGAATTTGATGTTTGGGTAAT 6754  
Qy 1007 CACCTCAAGATACACTATTGGCAATTTGACAAATATGTCACCTCGTATGAAACAATAT 1066  
Db 6755 AAATAAATGATATAGAGTTTAAATTAAGTTGTTAGTGGTTTAAATAGAGT 6814  
Qy 1067 AGTAGATGTTGTTTCTGCTCAAAATAAATAATTTTCTGTT 1109  
Db 6815 AGTAATTTGTTATTTTAGAGGATGTTGAGAAATGTTAGGGT 6857

## RESULT 2

US-08-998-416-288  
; Sequence 288, Application US/08998416  
; Patent No. 6239264

## GENERAL INFORMATION:

APPLICANT: Philippsen, Peter  
APPLICANT: Pohlmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine  
APPLICANT: Wendland, Jurgin  
APPLICANT: Knechtel, Philipp  
APPLICANT: Reibschung, Corinne  
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPIL  
TITLE OF INVENTION: AND USES THEREOF  
NUMBER OF SEQUENCES: 1152  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6239264artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: No. 6239264th Carolina  
COUNTRY: USA  
ZIP: 27709

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meliss, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 288:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 837 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE: PAG1241RP  
ORGANISM: US-08-998-416-288

Query Match 1.9%; Score 68.6; DB 3; Length 837;  
Best Local Similarity 47.6%; Pred. No. 2.7e-08;  
Matches 265; Conservative 0; Mismatches 289; Indels 3; Gaps 2;

Qy 524 TTAATTCGTTACTGCTCATATTCATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 583  
Db 154 TTAATTCATCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 213  
Qy 584 GTTATTCGATGAATATTCGAAATTA--TCACATAATAATTAATTAATTAATTAATTAATTC 641  
Db 214 TTAGTCTATGTTCAAAATTTTAAATTTAGTTTAAATTAATTAATTAATTAATTAATTTCT 273  
Qy 642 GTATTGTTATCAACTTTTCATGCTCTATACATGTAATTAATTAATTAATTAATTAATTAATTC 701  
Db 274 TTAAT-AAAATTTAAATAGATTAATCAATAATAATTAATTAATTAATTAATTAATTTGTTTAA 332  
Qy 702 AAGTAATTTGCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 761  
Db 333 AAATAATAATTTTATTAATTAATAAAGATTTAAATTTTAAATTAATTAATTAATTAATTAATTT 392  
Qy 762 TTTTATGTTTGTTCATTTTAAATGAGGTTGAGCTTTGTCATTCATTAATTTTATGATG 821  
Db 393 TTATTAATAATCAATTTTATAAATAATTAATTAATTAATTAATTAATTAATTAATTTTATA 452  
Qy 822 ACAACATCTTTGATGAAGTATTTAAGATATTTGTTAATGCAATGAGGGGTTTCCGTGTTATTT 881  
Db 453 AGAATTTATTAATAAATTAATTTAACTTTAACTTTAACTTTAACTTTAACTTTAACTTTATTT 512  
Qy 882 TTTATTAATAATCATATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 941  
Db 513 AATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 572  
Qy 942 CTTTAAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1001  
Db 573 ATTTTATCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 632  
Qy 1002 TAAATTCACCTCAAGATACACTATTGGCAATTTGCAAAATTCACAAATATGTCACCTTCGTATGAAACA 1061  
Db 633 TAAAGAGATCCGACCATTAATTTGTTTATGAGACAAATGCTTTAGCCCATAGCTATA 692  
Qy 1062 ATATTAGTAGATGTTGT 1078  
Db 693 TAGTTTGACTATCAATTT 709

## RESULT 3

US-08-487-826B-13/c  
; Sequence 13, Application US/08487826B  
; Patent No. 5993827

## GENERAL INFORMATION:

APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:



	Query Match	1.8%;	Score 65.2;	DB 4;	Length 6040;
	Best Local Similarity	46.0%;	Pred. No. 5.3e-07;		
	Matches 261;	Conservative	0;	Mismatches 303;	Indels 4; Gaps 1;
Qy	548	ATTAAATTCAC	TATTAAAT	TAACACAGTTC	TAAAAAGCGCTGTTTATTGGATGAATATTCGAAAT
Db	1395	ATTAGTGTGGTTTT	TAAGAATGATAT	TTTTTTATTTTTTTTGTTTAAAGTTTTTATTTAAGT	

Qy	608	TATCACAATAAATAATGGATGCGTATTATTACTGCTGTAATGGATCAACTTTCTATGCTCTA	667
Db	1455	AAGTAATTTAAATATTATTTAAATATAAAGTTTTTGGTATGTTATAAAAGTTTTTTTAAAAAT	1514
Qy	668	TACATGTAATATATTTTCGGAGTTAGACCTTAATTCGAAGGTAATTT---GTCTATTTAAT	723
Db	1515	TGGTTTAAATTTTTTAGTTTTATTTTATTATTAGTTATTATATTGATTTTTTTTTTATT	1574
Qy	724	ATTATCTGAATAATATGTAATCGATGCTTTTGGGTTATTTTTATGTTTGGTTTCATTTTT	783
Db	1575	TTTGTATTATTAAGGTTTTATGTAATATTGAATTTATGTTTTTTTGTTTTTTATAATTGT	1634
Qy	784	AATGACGGTGAGCTTGTCGATTCATATTTTTTTATGATGACAACATCTTTTGATCAAGTATT	843
Db	1635	TATTTTTTTTATTACGGGATAGAGATGTTGTTTTATTTTTTAGTTTAATGATTTTTTTTT	1694
Qy	844	TAAGATATTGTTTAATGCATGAGGGGTTGCGTGATTTTTTATATTAATTAATCAATAAAA	903
Db	1695	TAATATTTTTTAAAGTATAGGTGAGAGAGAAATATTTTTTTTTTTGGGGATATTTTTTAAG	1754
Qy	904	TCAACAATATATGTTATTTTGTCTTTTTATAGTCTCTTTTAAAGAGGTAGGATGACC	963
Db	1755	TAGTTTTTAGTTAGTATTATTTTTTTTTTATGTTGTTTAATGATTTTTTATATATTTTGT	1814
Qy	964	TAAAGTCCGCTAAATPATGGCTAAATGCCATTGCTATAATTCACCTCAAGATACACT	1023
Db	1815	TATAAGAGGAGAGAAAATATTTTATATTTATTTTTTGGAGATTTAGTTTAAATATCGAAAT	1874
Qy	1024	ATTGGCAAAATTGACAAATATGTCACCTTCGTATGAAACAATATTAGTAGATGTTGTTTTTG	1083
Db	1875	GTGGGTAATATATTTTTGTTTGTAGAGTTTGTGAAAAATATTATTAGTGTAATGAAGAAAAG	1934
Qy	1084	CTGCAAAAAATAAAAAATTTTCTGGTTGA	1111
Db	1935	ATATGAAATAAATTTTTTTTAGAGTTAA	1962

[illegible]



Matches 169; Conservative 0; Mismatches 156; Indels 3; Gaps 1;  
QY 3258 AGCGGTAGTAAGAACCGGTGTCAGCATCACCTATGACGTGGAACGAATGGCGCGTAA 3317  
Db 84 ACCGGTGGCCCGCGCGGTGCGAGATCACCTACGACGTGAGATCGGCAACGCCATCG 143  
QY 3318 AGACGAAGAGCTGCCGTTTGTGTTGGCGTCACTTGGCGACTTTTCAGGACACAAACCCAG 3377  
Db 144 AGAAGAGGAGCTGCCGCTGGTGGCATCTCGCCGACCTCTCCG---CAAGCCGG 200  
QY 3378 AATCAGAAAGTGTATTAGAGAGGAGAGTTCAGGGTATCGATAAGACAACTTCG 3437  
Db 201 ATACCCCGCGGCGAAGCTGGTGAACGGCGTTTCGTGCAATCGACCGCGACAATCCA 260  
QY 3438 ATACAGTGTAGGGGCAAAATCACCCGCGTCTTTCGTACAGGTTGTATAACAAGCTTGTCTA 3497  
Db 261 ACGGATCTCTCTCCATCTCTCCCGCGCCACCTCGAGGTCGACACACATCAGCG 320  
QY 3498 ATGATGATCCGAGTTTGAAGTGAAGTTCAGCTTCGTTGATGAAAGATTTCCACCCAG 3557  
Db 321 GCGACGACAGCAAGCTGACGTGAGCTGGCTTCAACCATCGAGGACTTCGACCCGG 380  
QY 3558 AGAAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3585  
Db 381 TCAACCTGGTGAACGAGGCTGCTCCGTT 408

RESULT 8  
US-10-204-708-63  
; Sequence 63, Application US/10204708  
; Patent No. 6677731  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication  
; FILE REFERENCE: 5013.1012  
; CURRENT APPLICATION NUMBER: US/10/204,708  
; CURRENT FILING DATE: 2003-05-06  
; PRIOR APPLICATION NUMBER: PCT/EP01/03971  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: DE 10019058.8  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: DE 10019173.8  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 98  
; SEQ ID NO 63  
; LENGTH: 5562  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-204-708-63

Query Match 1.7%; Score 60.4; DB 4; Length 5562;  
Best Local Similarity 48.0%; Pred. No. 9.5e-06;  
Matches 172; Conservative 0; Mismatches 186; Indels 0; Gaps 0;  
QY 583 TGTTTATGATGAATTCGAAATATACATAATTAATGATGCTATTATTACTTGCTG 642  
Db 1351 TTGTTTGGGGTGAATTTTTTTTTTTTTTTTTTTTTTTTATATATATATATATATTTT 1410  
QY 643 TATGGTATCAACTTTCATCTCTATACATGTAATATATATTCGAGTTAGACCTTAATCA 702  
Db 1411 TTTTTCGTTAGGAATGATTTTTTTTTTTGATATATATATATATTTTTTTTATTTAGTT 1470  
QY 703 AGGTAATTTCTATTAAT 762

Db 1471 TAGTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTATTTATTTTATTTTGTAGTAGATTTT 1530  
QY 763 TTTTATGTTTCTTCATTTTAAATGACCGTGTGTCATTGTCATATTTTTTATGATGA 822  
Db 1531 TTTTATTTATTTTGTGTTTTTATAGTGTTTTTTATTTTGTGTTTAAAGAA 1590  
QY 823 CAACATCTTGTGATGAAGTATTTAAGATATCTTAATGATGAGGGTTTGGTGTATTTT 882  
Db 1591 AGTTTATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTAG 1650  
QY 883 TTATTTTAAATCAATAATAAATCAACATATATCTTATTTTGTCTCTTTTATAGTGT 940  
Db 1651 AATTTTGTGATAATAATGATTTTATATATATTTATTTTGTGTTTTTATTTTAT 1708

RESULT 9  
US-10-204-708-20  
; Sequence 20, Application US/10204708  
; Patent No. 6677731  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication  
; FILE REFERENCE: 5013.1012  
; CURRENT APPLICATION NUMBER: US/10/204,708  
; CURRENT FILING DATE: 2003-05-06  
; PRIOR APPLICATION NUMBER: PCT/EP01/03971  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: DE 10019058.8  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: DE 10019173.8  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 98  
; SEQ ID NO 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-204-708-20

Query Match 1.7%; Score 60; DB 4; Length 6866;  
Best Local Similarity 44.7%; Pred. No. 1.3e-05;  
Matches 287; Conservative 0; Mismatches 345; Indels 10; Gaps 1;  
QY 495 CTTAATGCTGAGGAGCGGTTTGGTGGCTTTAATTCGTACTGCTCATATCAATTAAT 554  
Db 5366 CGTTATGTTGGTAGGTGTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTT 5425  
QY 555 CACTATTAAATAAACAGTTCTAAAAGGCTGTTTATTTGGATGAATATTCGAAATATCACA 614  
Db 5426 TTTTAAAGTGTGAGATTATAGGTATGATTTATTTTATTTTAAAGTTGTTTTTA 5485  
QY 615 TAAATTAATGATGCTATTTACTTCTGCTGATGTTGATCAACTTTCATGCTCTATACATG 674  
Db 5486 TATTTTATGCTTTTTCGATGTTTTTTTGTGTTTAAATTTTAAATTTTAAATTTTAAAT 5545  
QY 675 AATATTTTCGAGTTAGACCTTAATTTCAAGGTAATTTCTATTTTAAATTTATTCGTAAT 734  
Db 5546 TATTTTATGAAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTGT 5605  
QY 735 AATATGTAATCGATGCTTTGTTGTTTATTTATTTTATTTTATTTTATTTTATTTTAT 794  
Db 5606 ATTTTGTGTTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTAT 5665  
QY 795 GCTTGTGATTCATATTTTATGATGACACACTTTTATGATGAGTATTTTATGATTTTAT 854

Db	5666	GTTTATTTTTTATTTATTTTATAGATTAGATATATGTAGATTTTTTATATATTTA	5725
Qy	855	TAATGCATGAGGGTTCGCGTATTTTTTATATTAATCAATCAACAAATATA	914
Db	5726	TTTATAGATATATTTAGATATAAATATATATTTTTTATTTAAATTTAGATATAGATAGA	5785
Qy	915	TGTTATTTTGTCGCTTTTTTATAGTGTCTTTTAAAGAGGTAGGATGACCTAAAGGTCGCC	974
Db	5786	TAGAAATTTATAGTTGGTTTTTTGCGGGTTTTGTATTTGCGAATTTAATTAATTTGCGTT	5845
Qy	975	TAATATGCGGPAATTCGCCATGCTATAATTCACCTCAAAGATACACATTGGCAAAAT	1034
Db	5846	TGAAATATCGAAAA-----GAAAAATTAATTAATAATAATATAGTAATAA	5895
Qy	1035	GACAAATATGTCACATTCGTATGAACAATATTAGTAGATGTCGTTTTTCTGCAAAAATA	1094
Db	5896	ATAAATTTATAAAATTAGAAAAGTTAAATTTATTTGAGGTTTTTTTTTAATATTTTATTTA	5955
Qy	1095	AAAAATTTTTTCGGTTGAAATAACTCAAGGCCTCTAGCGTTTT	1136
Db	5956	AAATTTGATATTTTGGTTTTTAAATTAATAATAATAGCGATTT	5997

RESULT 10  
US-09-790-988-1

Query Match	1.7%;	Score 59.8;	DB 4;	Length 640681;	
Best Local Similarity	44.1%;	Pred. No. 0.00013;			
Matches	Conservative	0;	Mismatches 317;	Indels 0;	Gaps 0;
QY	515	TTTGGTGCTTTAAATTCGTTACTGCTCATATTCATTTAAATTCACCTATTAATTAACAGATTC	574		
Db	325026	TTTGATATTAAATAATAATTTTTTGTGTTTTTAAAAATTTATTTTTTGTGTAATTTTCA	325085		
QY	575	TAAAGGCGTGTATTGTCATGAATTCGAAATTTACATAATACACATAATAATTCATGTCGATTATT	634		
Db	325086	TAGACGTAATTTCTTTAAATAAGTAATAATTTTTTAATTAATTAACATAATTTTAT	325145		
QY	635	ACTTGCTGTAATGGTATCAACTTCATGCTCTATACATGTAATATATTTTCGAGTTAGACC	694		
Db	325146	TCATATTATATTTTTTATAAAATTTAAAAAATATTTCTTAAAAATAATTTTAAATGTGTAT	325205		
QY	695	TTAATTCAGGTAATTTGCTATTTTAATTAATATCTGAATAATATGTAATCGATTCGTTT	754		
Db	325206	TTAAAAAATAATAATAATAATTTTATTTTACATTTTATTATATATAAAATTTATCT	325265		
QY	755	GTGGTATTTTTATGTTTGTTCATTTTTTAATGACGGTGAGCTGTGCAATTCATATTTTT	814		
Db	325266	ATAAATATTTTTATATTTTAAATAAATAACATTTTATATAAAATAAATACTGATATTTTT	325325		
QY	815	TATGATGACACATCTTTGATGAAGTATTTAAGATATTTGTTAATGCAATGAGGGTTTGGC	874		
Db	325326	TCTTTTATTTTTTTTAAATTACTATAATGATAATTTAGTAGTATATAATAAATAAGTAAAAA	325385		

```

RESULT 11
US-09-252-991A-752
; Sequence 752, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 752
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-752

```

	Query Match	1.7%	Score 59.4;	DB 4;	Length 600;
	Best Local Similarity	56.1%;	Pred. No. 6.1e-06;		
	Matches 133;	Conservative	0;	Mismatches 101;	Indels 3; Gaps 1;
QY	3275	CGTGTGAGCATCACCTATGACGTTGTAACGATGGCGCGCTAAAGACGAAAGAGCTGCCG	3334		
Db	133	CGCGTCAGATCGAGTACGACGTGGAGTTGTACGGTGGCGAGAGAGAGTCCAGCTGCC	192		
QY	3335	TTTGTGTTGGGTCATTGGCGACTTTTCAGGACACAAACCGAATCAGAAAAGTTGAT	3394		
Db	193	TTGCTCATGGCGTCATGCGCGACCTCGCGG---CAAGCCCGCAACCCAGCGCGG	249		
QY	3395	TTAAGAGCGAGATTACGGGTATCGATTAAGACAACTTCGATACAGTATGGGGCAA	3454		
Db	250	GTCGCCGACGGCAAGTTCCTGGAGATCGACGTGGACAACTTCGATGCCGGCTGAAGCG	309		
QY	3455	ATTACCCCGCTTTTCGTACAGGGTTGATTAACAAGCTTGCTAATGATGATAGCCAG	3511		
Db	310	ATTACGCGCGGTGCGCTTCACGTTACCGACGTCGTGACCGCGGAAGCAACCTG	356		

RESULT 12  
US-09-252-991A-723  
; Sequence 723, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788

;; PRIOR FILING DATE: 1998-02-18  
;; PRIOR APPLICATION NUMBER: US 60/094,190  
;; PRIOR FILING DATE: 1998-07-27  
;; NUMBER OF SEQ ID NOS: 33142  
;; SEQ ID NO 723  
;; LENGTH: 681  
;; TYPE: DNA  
;; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-723

Query Match 1.7%; Score 59.4; DB 4; Length 681;  
Best Local Similarity 56.1%; Pred. No. 6.5e-06;  
Matches 133; Conservative 0; Mismatches 101; Indels 3; Gaps 1;  
QY 3275 CGTGTACGATCACCTATGACGTTGAAACGAATGGCCGCTAAAGACGAAGAGCTGCCG 3334  
DB 140 CGGTGCAGATCAGTACGACGCTGGAGTTGTACGGTCCGAGAGAGGTCCAGCTGCC 199  
QY 3335 TTGTTGTTGGGTCATGGCGACTTTTCAGGACACAAACAGAAATCAGAAAAAGTTGAT 3394  
DB 200 TTGTCATGGCGTCATGGCGACCTCGCCG---CAAGCCCGCCGAAACCCAGGGCGG 256  
QY 3395 TTGAAGAGCGAGAGTTACGGGTATCGATAAAGACAACTTCGATACAGTGTGGGGCAA 3454  
DB 257 GTCCGACCGCAAGTTCTGGAGATCGAGTGGACAACTTCGATGCCCGGCTGAAGGG 316  
QY 3455 ATTACCCGGGCTCTTCGTACAAGGTTGATAACAAGCTTGTCTAATGATGATGCCAG 3511  
DB 317 ATGAAGCCGGGTGGCTTCAACGTACCGAAGCTGTGACCGCGGAAGCAACCTG 373

RESULT 13  
US-09-252-991A-658/c  
;; Sequence 658, Application US/09252991A  
;; Patent No. 6551795  
;; GENERAL INFORMATION:  
;; APPLICANT: Marc J. Rubenfield et al.  
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
;; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
;; FILE REFERENCE: 107196.136  
;; CURRENT APPLICATION NUMBER: US/09/252,991A  
;; PRIOR FILING DATE: 1999-02-18  
;; PRIOR APPLICATION NUMBER: US 60/074,788  
;; PRIOR FILING DATE: 1998-02-18  
;; PRIOR APPLICATION NUMBER: US 60/094,190  
;; PRIOR FILING DATE: 1998-07-27  
;; NUMBER OF SEQ ID NOS: 33142  
;; SEQ ID NO 658  
;; LENGTH: 2019  
;; TYPE: DNA  
;; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-658

Query Match 1.7%; Score 59.4; DB 4; Length 2019;  
Best Local Similarity 56.1%; Pred. No. 1.1e-05;  
Matches 133; Conservative 0; Mismatches 101; Indels 3; Gaps 1;  
QY 3275 CGTGTACGATCACCTATGACGTTGAAACGAATGGCCGCTAAAGACGAAGAGCTGCCG 3334  
DB 1935 CGGTGCAGATCAGTACGACGCTGGAGTTGTACGGTCCGAGAGAGGTCCAGCTGCC 1876  
QY 3335 TTGTTGTTGGGCTCATTTGGCGACTTTTCAGGACACAAACCAATCAGAAAAGTTGAT 3394  
DB 1875 TTGTCATGGGGTCATGGCGACCTCGCCG---CAAGCCCGCCGAAACCCAGGGCGG 1819  
QY 3395 TTGAAGAGCGAGAGTTACGGGTATCGATAAAGACAACTTCGATACAGTGTGGGGCAA 3454  
DB 1818 GTCCGACCGCAAGTTCTGGAGATCGAGTGGACAACTTCGATGCCCGGCTGAAGGG 1759  
QY 3455 ATTACCCGGGCTCTTCGTACAGGTTGATAACAAGCTTGTCTAATGATGATGCCAG 3511  
DB 1758 ATGAAGCCGGGTGGCTTCAACGTACCGAAGCTGTGACCGCGGAAGCAACCTG 1702

RESULT 14  
US-08-883-795A-36  
;; Sequence 36, Application US/08883795A  
;; Patent No. 5985607  
;; GENERAL INFORMATION:  
;; APPLICANT: Delcuve, Genevieve  
;; APPLICANT: Awang, Gregor  
;; TITLE OF INVENTION: Recombinant DNA Molecules and Expression  
;; TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator  
;; NUMBER OF SEQUENCES: 39  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: BERESKIN & PARR  
;; STREET: 40 King Street West  
;; CITY: Toronto  
;; STATE: Ontario  
;; COUNTRY: Canada  
;; ZIP: M5H 3Y2  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/883,795A  
;; FILING DATE: 27-JUN-1997  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Gravelle, Micheline  
;; REGISTRATION NUMBER: 40,261  
;; REFERENCE/DOCKET NUMBER: 7841-062  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (416) 364-7311  
;; TELEFAX: (416) 361-1398  
;; INFORMATION FOR SEQ ID NO: 36:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 665 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: cDNA  
;; ORIGINAL SOURCE:  
;; ORGANISM: Homo sapiens  
;; IMMEDIATE SOURCE:  
;; CLONE: Rh 32  
;; US-08-883-795A-36

Query Match 1.6%; Score 59.2; DB 2; Length 665;  
Best Local Similarity 45.8%; Pred. No. 7.2e-06;  
Matches 242; Conservative 0; Mismatches 283; Indels 3; Gaps 1;  
QY 515 TTGCGCTTTAAATTCGTTACTGCTCATATTCATTAATTCACATTAATTCACATTAATTAACAGTTC 574  
DB 112 TTTTATAATTAATAATTTATAATTAATAATTTATAATTAATAATTTATAATTAATAATA 171  
QY 575 TAAAGCGCTTTATTGGATGAATTCGAAATATTCACATAATTAATTAATTAATTAATTAATTAAT 634  
DB 172 TTTTATAATTAATAATTTATAATTAATAATTTATAATTAATAATTTATAATTAATAATA 231  
QY 635 ACTTGCTGTTATGGTATCAACTTTCATGCTCTATACATGTAATTAATTTCCGAGTTAGACC 694  
DB 232 TTTTATAATTAATAATTTATAATTAATAATTTATAATTAATAATTTATAATTAATAATA 291  
QY 695 TTAATCAAGGTAATTTCTCTATTATAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 754  
DB 292 TTTTATAATTAATAATTTATAATTAATAATTTATAATTAATAATTTATAATTAATAATA 351  
QY 755 GTGGTATTATTTA---TGTGTTTTCATTTTAAATGACGGTGAGCTTGTGCAATTCATATT 811  
DB 352 TTTTATAATTAATAATTTATAATTAATAATTTATAATTAATAATTTATAATTAATAATA 411  
QY 812 TTTTATGATGACAAACATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGTTT 871

Db 10958 AAAGGAATTTATGTTTAAACGTAATTTATATATTTTATTTATTTT 11005

Search completed: July 2, 2004, 22:59:43  
Job time : 244 secs

Db 412 TTTTATAATTAATGTTTATAATTAACATATTTTATAATTAACATATTTTATAAAGTATTT 471  
QY 872 GCGGTATTTTATATTAATCAATAAAATCAACAATATATGTTATTTTGTCTTT 931  
Db 472 ATAATTAACATATTTTATAATTAAGTATTTTATAATTAACATATTTTATAATTAAGTATTT 531  
QY 932 TTATAGTGTCTTTTAAAGAGGTAGGATGACCTAAAGTCGCCCTAAATATGGCGTAAAT 991  
Db 532 ATAATTAACATATTTTATAATTAATCAATTTTATAAATAGTTTAAAAAGACGAGGAAAAAAT 591  
QY 992 GCCATTGCTATAATTCACCTCAAGATACACTATTGGCAATTCACAA 1039  
Db 592 AAAAGACGAGGTATTGATCTCAGGAATTGTAATTCGCAAGTGAGAA 639

## RESULT 15

US-10-204-708-55  
; Sequence 55, Application US/10204708  
; Patent No. 6677731  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication  
; TITLE OF INVENTION: by Assessing DNA Methylation  
; FILE REFERENCE: 5013.1012  
; CURRENT APPLICATION NUMBER: US/10/204,708  
; CURRENT FILING DATE: 2003-05-06  
; PRIOR APPLICATION NUMBER: PCT/EP01/03971  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: DE 10019058.8  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: DE 10019173.8  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 98  
; SEQ ID NO 55  
; LENGTH: 11015  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-204-708-55

Query Match 1.6%; Score 57.6; DB 4; Length 11015;  
Best Local Similarity 49.7%; Pred. No. 7.2e-05;  
Matches 173; Conservative 0; Mismatches 174; Indels 1; Gaps 1;  
QY 595 TTTATTGATCAATATTCGAAATTCACATAAATGATGCTATTATTACTTGTCTGTA 644  
Db 10659 TTTTAGAGTGGTTAAAGTATTTTATATTTTATGATGAATGTTTAAATTTTATATTT 10718  
QY 645 TTGATATCAACTTTCATGCTCTATACATGTAATATATTTTCGAGTAGACCTTAATTCAG 704  
Db 10719 TTAATAATTTTGTATTTTATTTTGTGTTTTTATTTTATTTTATTTTATTTTAAAT 10778  
QY 705 GTAATTTGCTATTAATTAATATATCTGAATAATATGTAATGATGCTTTTGTGTTATTT 764  
Db 10779 GGGTATGGAGTGGTATTTTATTTATTTGTTTGGTTTGGTTTAAAGTAAATAT 10838  
QY 765 TTATGTTTGTTCATTTTTTAAATGACGGTGGCTTTGTGCATTCATATTTTTTATGACACA 824  
Db 10839 TGAGTATTTTTTATACT-TTGTTGTTTATTTTTTATTTTTTTTGGAGTAAATGTTGTT 10897  
QY 825 ACATCTTTGATGAAGTATTTAAGATATTTGTTAATGCAATGAGGGGTTTGGCGTATTTTT 884  
Db 10898 ATATTTTGGTATTATATAATTAAGAAATTTGTTTAAATTTAGGATTAGAAGAAATATAT 10957  
QY 885 ATATTAAATCATATAAATAATCAACAATATATGTTATTTTGTCTTTT 932







Qy 753 TTGGGTTATTTTATGTTTTCATTTTAAAGACGGTGTGCAATCATATTT 812  
Db 564488 GAAATTTAAATTTATGTTTTTTTTTTTATAGATTATGTTTTAAATGTTAAAGTTAAGAA 564547  
Qy 813 TTTATGATGACACATCTTTTGATGAAGTATTTAAGATATTTGTTAATGCAATGAGGGGTTTG 872  
Db 564548 TTTTTCGTTTAGATTATAT-TTAAATATTTTTTATTTTTTAAATTTAAGAGTTTTA 564606  
Qy 973 CGTGTATTTTTATATTAATCAATAATAAATCAACAATATATGTTTATTT 922  
Db 564607 TGTTTTATATTTTATATAATTTATTTGTTTATTTTATTTTAAATTAAT 564656

RESULT 4  
US-10-221-714A-514  
; Sequence 514, Application US/10221714A  
; Publication No. US20040048254A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with  
; TITLE OF INVENTION: tumor suppressor genes and oncogenes  
; FILE REFERENCE: 5013.1005  
; CURRENT APPLICATION NUMBER: US/10/221.714A  
; CURRENT FILING DATE: 2003-01-21  
; PRIOR APPLICATION NUMBER: PCT/EP01/02955  
; PRIOR FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: DE 10013847.0  
; PRIOR FILING DATE: 2000-03-15  
; PRIOR APPLICATION NUMBER: DE 10019058.8  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: DE 10019173.8  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 540  
; SEQ ID NO 514  
; LENGTH: 61020  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-221-714A-514

Query Match 2.0%; Score 71.4; DB 13; Length 61020;  
Best Local Similarity 48.2%; Pred. No. 3.6e-05;  
Matches 201; Conservative 0; Mismatches 216; Indels 0; Gaps 0;  
Qy 517 TGGTGCTTTAATTCGTTACGTCATATTCATTAATTCATTAATAAAGACGTTCTA 576  
Db 11783 TGTTATATAATTTGTTATATTTTATATTTATTTTATATATATATATTTT 11842  
Qy 577 AAGCGTGTTTATGGAGTAATATCGAANTATCAATAAATTAATGCTATATTTAC 636  
Db 11843 GGAAATATATAGGTTTTTTTAAAGTTTAAATAATTTTAAATTAATTTATTTT 11902  
Qy 637 TTGCTGATGGATCAACCTTCATGCTCATACATGTAATATATTCGAGTTAGACCTT 696  
Db 11903 TAAATTTTTTTTGTAGTATTTTATTTAAGTAAATATATATTTTAAATTTAT 11962  
Qy 697 AATTCAGGTAATTTGTCTATTTAATTTATCTGAATATATGTAATCGATGCTTTGT 756  
Db 11963 GTTTGTCAGAAGTTTGGGATTTTTTTTTTTTTTTTTTTTATTCGTTATTTGTT 12022  
Qy 757 GGTATTTTTATGTTGTTTCATTTTTTAATGACGGTGGCTTGTGCAATCATATTTT 816  
Db 12023 TGTATAAATTTATTTTATATTTTGTAACTTATTTTGTTTTATATTTATTTT 12082  
Qy 817 TGATGACACATCTTTGATGAAGTATTTAAGATATTTGTTAATGATGAGGGGTTTGGGTG 876

Db 12083 TGTTAGTTTTTATTTTTTTTTTTTTTTTAAATTTTAGAGTAATTTATATTTTTTTT 12142  
Qy 877 TATTTTATATTAATAATCAATAATAAATCAACAATATATGTTATTTTGTGCTTTTT 933  
Db 12143 TGTTTTATATATTTTTTTTTTTTAAAGTTTATTTTCGTTGATATTTGTTATTTAT 12199  
RESULT 5  
US-10-221-714A-158  
; Sequence 158, Application US/10221714A  
; Publication No. US20040048254A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with  
; TITLE OF INVENTION: tumor suppressor genes and oncogenes  
; FILE REFERENCE: 5013.1005  
; CURRENT APPLICATION NUMBER: US/10/221.714A  
; CURRENT FILING DATE: 2003-01-21  
; PRIOR APPLICATION NUMBER: PCT/EP01/02955  
; PRIOR FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: DE 10013847.0  
; PRIOR FILING DATE: 2000-03-15  
; PRIOR APPLICATION NUMBER: DE 10019058.8  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: DE 10019173.8  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 540  
; SEQ ID NO 158  
; LENGTH: 8093  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-221-714A-158

Query Match 1.9%; Score 69.4; DB 13; Length 8093;  
Best Local Similarity 47.5%; Pred. No. 3e-05;  
Matches 277; Conservative 0; Mismatches 296; Indels 10; Gaps 2;  
Qy 527 ATTCGTTACTGCTCATATTCATTAATTCACATTAATAAAGACGTTCTAAAAGCTGTT 586  
Db 6285 ATTGTTGTTTATTTTAAABAATTTGTTTAAAGTAATATTTGTTTAAATGATATA 6344  
Qy 587 TATTCGATGAATATTCGAAATATCACATAATATGATGCTATTTACTTGTGTTATT 646  
Db 6345 GTATTAGATTTTGTAGATGTAGAAATGGATTTATTTTAAATTTGGAATTTGCTATA 6404  
Qy 647 GGTATCACTTTTCATGCTCTATACATGTAATATATTTTCGAGTTAGACCTTAATCAAGT 706  
Db 6405 TATTTTATGATGATAGATAGTATATAGTAGAAATATTTTAAAGTAGTTTTTATATAGAT 6464  
Qy 707 AATTTGCTATTTAATTTATCTGAATATATGTAATGATGCTTTGTTGTTATTTTT 766  
Db 6465 TGTAGTAATTTGTATTTTATTAAGATAATTTGTTTGTGTTAAATAGTAATTTTAA 6524  
Qy 767 ATGTTTGTTCATTTTAAATGACGGTGTGCTTGCATTCATATTTTTTATGATGACAC 826  
Db 6525 ATTTTGTGTTATATGAAGAAGTAATTTTAAAGTTTATTTATGTAATTTAATTAATA 6584  
Qy 827 ATCTTTGATGAATTTAAGATATTTGTAATGATGAGGGGTTTTCGTTATTTTTAT 886  
Db 6585 GGATTTAATTTATATTTATAGATTTTTTAAAGTATTATATATTTT-AAAAATTTTCTT 6642  
Qy 887 ATTAATCATATAAATAAATCAACAATATATGTTATTTTGTGCTCTTTTATAGTGTCTTT 946  
Db 6643 TTAAAGTTTGTGTTTAAATTTATTTGAATATTTTAAATTTGTTGTTATTTATTAATAATTT 6702

```
QY 947 AAGAGAGTGGATGACCTAAGCGTGGCTAAATGATGCGGTAATTTGCCATGCTATAATT 1006
DB 6703 AAGGAATTTAGGTTATGATTAG-----AATTTAGGTGGAATTTGATGTTGGGTAATT 6754

QY 1007 CACCTCAAGATCACACTATGGCAAAATGCAAAATATGTCACATCGTATGAAACAATATT 1066
DB 6755 AAAATAATGTTAAGAGTTTAAATAATTAAGTTGTTAGTGGTTTAAATTAGAGGT 6814

QY 1067 AGTAGATGTTTCTGTCGCAAAATAAATAATTTTCTGGTT 1109
DB 6815 AGTAATTTGTTATTTAGAGGATGTTGAGAAATGCTAGGGGT 6857

RESULT 6
US-10-172-086-26
; Sequence 26, Application US/10172086
; Publication No. US20030113750A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Method and nucleic acids for the differentiation
; TITLE OF INVENTION: of prostate tumors
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/172.086
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 26
; LENGTH: 8093
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-172-086-26

Query Match 1.9%; Score 59.4; DB 15; Length 8093;
Best Local Similarity 47.5%; Pred. No. 3e-05;
Matches 277; Conservative 0; Mismatches 296; Indels 10; Gaps 2;

QY 527 ATTCGTTACTGCTCATATTCAAATTAATCACTATTAATAAACAAGTCTTAAAGGCTGTT 586
DB 6285 ATTTGTTGTTTATTTTAAATAATTTGTTTAAAGTAATATTTGTTTAAATGATATA 6344

QY 587 TATTCGATGAATATTCGAATATTCACATAATGATGCTATTAATTTACTTCTGTAATT 646
DB 6345 GTATTAGATTTTGTAGATGTTAGAAATGGAATTTTAAATTTGGAATTTGCTGATA 6404

QY 647 GGTATCAACTTTCATGCTCTATACATGTAATATTTTCGAGTTAGACCTTAATTCAGGT 706
DB 6405 TATTTTATGTAAGATAGTATATAGTAGAAATATTTAAAGTAGTTTATTTATAGAT 6464

QY 707 AATTTGCTATTTAATTAATCTGAATAATATGTAATCGATGCTTTGTTGTTATTTT 766
DB 6465 TGTAATAATTTGATTTTATTAAGATAATTTGTTTGTCTTAAATAAGTAAATTTTAA 6524

QY 767 ATGTTTGTTCATTTTAAAGACGCTGAGCTTGTGCAATTCATATTTTATGATGACAAC 826
DB 6525 ATTTTGTGTTATTAAGAAAGGTAAATTTAAAGTTTATTAAGTAAATTAATAATA 6584

QY 827 ATCTTTGATGAAGTATTAAGATATTTTAATGATGAGGGTTTGGCTGATTTTTAT 886
DB 6585 GGAATTAATTTATTTATAGATTTTAAAGTATTAATAATTTTAAAGTATTAATAATA 6642

QY 887 ATTAATCATATAATAATCAATATATGTTATTTGTCCTTTTATAGTCTCTTTT 946
DB 6643 TTAAGTTTGTGTTTAAATATTTGAATATTTTAAATTTGTTGTTATTAATTTAGTAAATTT 6702

QY 947 AAAGAGTGGATGACCTAAGAGTGGCTAAATATGCGTAAATTTGCCATTCGTATAATT 1006
DB 6703 AAGGAATTTAGTTATGATTAG-----AATTTAGGTGGAATTTGTTGGGTAATT 6754

QY 1007 CACCTCAAGATCACACTATGGCAAAATGCAAAATATGTCACATCGTATGAAACAATATT 1066
DB 6755 AAAATAATGTTAAGAGTTTAAATAATTAAGTTGTTAGTGGTTTAAATTAGAGGT 6814
```

```
QY 1067 AGTAGATGTTTCTGTCGCAAAATAAATAATTTTCTGGTT 1109
DB 6815 AGTAATTTGTTATTTAGAGGATGTTGAGAAATGTTAGGGGT 6857

RESULT 7
US-10-204-708-32
; Sequence 32, Application US/10204708
; Publication No. US20030141852A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 32
; LENGTH: 8093
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-32
```

```
Query Match 1.9%; Score 69.4; DB 15; Length 8093;
Best Local Similarity 47.5%; Pred. No. 3e-05;
Matches 277; Conservative 0; Mismatches 296; Indels 10; Gaps 2;

QY 527 ATTCGTTACTGCTCATATTCAAATTAATCACTATTAATAAACAAGTCTTAAAGGCTGTT 586
DB 6285 ATTTGTTGTTTATTTTAAATAATTTGTTTAAAGTAATATTTGTTTAAATGATATA 6344

QY 587 TATTCGATGAATATTCGAATATTCACATAATGATGCTATTAATTTACTTCTGTAATT 646
DB 6345 GTATTAGATTTTGTAGATGTTAGAAATGGAATTTTAAATTTGGAATTTGCTGATA 6404

QY 647 GGTATCAACTTTCATGCTCTATACATGTAATATTTTCGAGTTAGACCTTAATTCAGGT 706
DB 6405 TATTTTATGTAAGATAGTATATAGTAGAAATATTTAAAGTAGTTTATTTATAGAT 6464

QY 707 AATTTGCTATTTAATTAATCTGAATAATATGTAATCGATGCTTTGTTGTTATTTT 766
DB 6465 TGTAATAATTTGATTTTATTAAGATAATTTGTTTGTCTTAAATAAGTAAATTTTAA 6524

QY 767 ATGTTTGTTCATTTTAAAGACGCTGAGCTTGTGCAATTCATATTTTATGATGACAAC 826
DB 6525 ATTTTGTGTTATTAAGAAAGGTAAATTTAAAGTTTATTAAGTAAATTAATAATA 6584

QY 827 ATCTTTGATGAAGTATTAAGATATTTTAATGATGAGGGTTTGGCTGATTTTTAT 886
DB 6585 GGAATTAATTTATTTATAGATTTTAAAGTATTAATAATTTTAAAGTATTAATAATA 6642

QY 887 ATTAATCATATAATAATCAATATATGTTATTTGTCCTTTTATAGTCTCTTTT 946
DB 6643 TTAAGTTTGTGTTTAAATATTTGAATATTTTAAATTTGTTGTTATTAATTTAGTAAATTT 6702

QY 947 AAAGAGTGGATGACCTAAGAGTGGCTAAATATGCGTAAATTTGCCATTCGTATAATT 1006
DB 6703 AAGGAATTTAGTTATGATTAG-----AATTTAGGTGGAATTTGTTGGGTAATT 6754
```





[illegible]

Db 4054 ATTAAGTAGCGTGATTTAGAGTTTGTTATTTTAAATAGTATT 4097

## RESULT 14

US-10-221-714A-27

; Sequence 27, Application US/10221714A

; Publication No. US20040049254A1

; GENERAL INFORMATION:

; APPLICANT: OLEK, Alexander

; APPLICANT: PIEPENBROCK, Christian

; APPLICANT: BERLIN, Kurt

; TITLE OF INVENTION: Diagnosis of Diseases Associated with

; TITLE OF INVENTION: tumor suppressor genes and oncogenes

; FILE REFERENCE: 5013.1005

; CURRENT APPLICATION NUMBER: US/10/221,714A

; CURRENT FILING DATE: 2003-01-21

; PRIOR APPLICATION NUMBER: PCT/EP01/02955

; PRIOR FILING DATE: 2001-03-15

; PRIOR APPLICATION NUMBER: DE 10013847.0

; PRIOR FILING DATE: 2000-03-15

; PRIOR APPLICATION NUMBER: DE 10019058.8

; PRIOR FILING DATE: 2000-04-06

; PRIOR APPLICATION NUMBER: DE 10019173.8

; PRIOR FILING DATE: 2000-04-07

; PRIOR APPLICATION NUMBER: DE 10032529.7

; PRIOR FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: DE 10043826.1

; PRIOR FILING DATE: 2000-09-01

; NUMBER OF SEQ ID NOS: 540

; SEQ ID NO 27

; LENGTH: 8666

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-221-714A-27

Query Match 1.9%; Score 66.6; DB 13; Length 8666;

Best Local Similarity 44.6%; Pred. No. 0.00015;

Matches 261; Conservative 0; Mismatches 324; Indels 0; Gaps 0;

QY 525 TAATTCGTACTGCTCATATTCAATTAATTCACATATAAATAACAGTCTTAAAGCGTG 584

Db 30 TAATTAGCGTGTATTTATGCTGTATGATTTTATTTTAGTTGGTGTATTTAGTTGGTT 89

QY 585 TTTATGGAATGAATTCGAAATATCACATAAATGATGCTATTATTACTTCTGCTGA 644

Db 90 TTAAGAATAAAATAAAGGTTTTTTTGTGAGTTTTATTTAAATTTATTTTATTT 149

QY 645 TTGGTATCAACTTTCATGCTCTATACATGTAATATTTTCGAGTTAGACCTTAATTCAG 704

Db 150 ATAAGAGGTATTTTAAATTTTACGATTTTAAATTTTAAATTTTAAATTTTATTTTA 209

QY 705 GTAATTCGTCTATTAAATTTATCTGAATAATGTAATGATGCTTTTGGTGTATTT 764

Db 210 TTTTGTAATAAGTAAATTTTATTTATAGAAGGAGTGTTTAAATTTTTCGTT 269

QY 765 TTAATGTTTGTTCATTTTATGACGGTGTGCTGTCATTCATATTTTATGATGACA 824

Db 270 TTGATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 329

QY 825 ACATCTTTGATGAAGTATTAAAGATATTGTTAAATGCAAGGGGTTTCGGTGTATTTT 884

Db 330 GTATTTTCAGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 389

QY 885 ATATTAATCATATAAATCAACAATATATGTTATTTTGTGCTTTTATAGTGTCTT 944

Db 390 ATTTGTTGTGATTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 449

QY 945 TTAAGAGGTAGGATGACCTAAAGGTCGCCCTAAATATGGCGTAAATTCGCAATTCATAA 1004

Db 450 GGTATTTGTTTTTTAGGTATATGCTGTGTTTTAGGTTTTTTTTTATTTTATTTTATTT 509

QY 1005 TTCACCTCAAGATACACTATTGGCAATTCGAAATATGCAAAATATGTCACCTCGTAAACAATA 1064

Db 510 TTTATGGTATAGGTTTTTGAATTTATTAATTAATTAAGTTATGAAGAGTAGTTAACGTAGTG 569

QY 1065 TTAGTAGATGTTGTTTTTGTCTGCAAAATAAATAATTTTCTGGTT 1109

Db 570 TTTTATTTTGTGTTATTTATTTATTTTAGTTTATTTTGGTT 614

## RESULT 15

US-10-311-455-369

; Sequence 369, Application US/10311455

; Publication No. US20030143606A1

; GENERAL INFORMATION:

; APPLICANT: OLEK, Alexander

; APPLICANT: PIEPENBROCK, Christian

; APPLICANT: BERLIN, Kurt

; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ

; TITLE OF INVENTION: cytosine methylation

; FILE REFERENCE: 5013.1014

; CURRENT APPLICATION NUMBER: US/10/311,455

; CURRENT FILING DATE: 2002-12-16

; PRIOR APPLICATION NUMBER: PCT/EP01/07537

; PRIOR FILING DATE: 2001-07-02

; PRIOR APPLICATION NUMBER: DE 10032529.7

; PRIOR FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: DE 10043826.1

; PRIOR FILING DATE: 2000-09-01

; NUMBER OF SEQ ID NOS: 2424

; SEQ ID NO 369

; LENGTH: 8666

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-311-455-369

Query Match 1.9%; Score 66.6; DB 15; Length 8666;

Best Local Similarity 44.6%; Pred. No. 0.00015;

Matches 261; Conservative 0; Mismatches 324; Indels 0; Gaps 0;

QY 525 TAATTCGTACTGCTCATATTCAATTAATTCACATAAATGATGCTATTATTACTTCTGCTGA 584

Db 30 TAATTAGGTTGTTTTATTGCTGTATGATTTTATTTTAGTTGGTGTATTTAGTTGGTT 89

QY 585 TTTATGGAATGAATTCGAAATATCACATAAATGATGCTATTATTACTTCTGCTGA 644

Db 90 TTAAGAATAAAATAAAGGTTTTTTTGTGAGTTTTATTTAAATTTATTTTATTTTATTT 149

QY 645 TTGGTATCAACTTTCATGCTCTATACATGTAATATTTTCGAGTTAGACCTTAATTCAG 704

Db 150 ATAAGAGGTATTTTAAATTTTACGATTTTAAATTTTAAATTTTAAATTTTATTTTA 209

QY 705 GTAATTCGTCTATTAAATTTATCTGAATAATGTAATGATGCTTCTGCTGTATTT 764

Db 210 TTTTGTAAATAAGTAAATTTTATTTATAGAAGGAGTGTTTAAATTTTTCGTT 269

QY 765 TTAATGTTTGTTCATTTTAAATGACGGTGTGCTGTCATTCATATTTTATGATGACA 824

Db 270 TTGATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 329

QY 825 ACATCTTTGATGAAGTATTAAAGATATTGTTAAATGCAATGAGGGGTTTCGGTGTATTTT 884

Db 330 GTATTTTCAGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 389

QY 885 ATATTAATCATATAAATCAACAATATATGTTATTTTGTGCTTTTATAGTGTCTT 944

Db 390 ATTTGTTGTGATTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 449

QY 945 TTAAGAGGTAGGATGACCTAAAGGTCGCCCTAAATATGGCGTAAATTCGCAATTCATAA 1004

Db 450 GGTATTTGTTTTTTAGGTATATGCTGTGTTTTAGGTTTTTTTTTATTTTATTTTATTTAT 509

Qy	1005	TTCCACCTCAAAGATACACTATTGGCAAAATTGACAAATATGTCACITTCGTATCGTATGAAACAATA	1064
Db	510	TTTATCGTATAGCGTTTTCGAATTATATTAATTAACTTAAGTAAGAAAGAGTAGTTAACGTAGTG	569
Qy	1065	TTAGTAGATGTTGTTTTTCCTGCAAAAATAAAAAATTTTTCTGGTT	1109
Db	570	TTTTTATTTTTTGTTATTATTATTATTATTAGTTTTTTTTTTGGTT	614

Search completed: July 3, 2004, 02:41:59  
Job time : 1507 secs

**THIS PAGE BLANK (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 2, 2004, 15:43:37 ; Search time 8676 Seconds

(without alignments)

12349.643 Million cell updates/sec

Title: US-10-780-347-1

Perfect score: 3588

Sequence: 1 gtcgacttattgcattgatg.....gatnaattgagccgcttaa 3588

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estro:\*  
7: em\_estro:\*  
8: em\_estro:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_est3:\*  
12: gb\_est4:\*  
13: gb\_est5:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_fod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrt:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	91	2.5	1101	29	CNS00EVL
c 2	82.6	2.3	1200	13	AL069706 Drosophila
c 3	82	2.3	1094	29	AL069706 Drosophila
c 4	80.2	2.2	1200	13	AL069706 Drosophila

c 5	79.2	2.2	999	13	AL069706 Drosophila
c 6	78.8	2.2	714	28	AL069706 Drosophila
c 7	77.4	2.2	1201	13	AL069706 Drosophila
c 8	77.2	2.2	712	13	AL069706 Drosophila
c 9	76.6	2.1	1101	29	CNS0039G
c 10	76.6	2.1	1101	29	CNS0039G
c 11	76.6	2.1	1201	13	AL069706 Drosophila
c 12	76.4	2.1	1201	13	AL069706 Drosophila
c 13	76.2	2.1	1200	13	AL069706 Drosophila
c 14	75.8	2.1	1200	13	AL069706 Drosophila
c 15	75.8	2.1	1201	13	AL069706 Drosophila
c 16	75.6	2.1	1128	13	AL069706 Drosophila
c 17	75.2	2.1	527	14	AL069706 Drosophila
c 18	75.2	2.1	781	29	CNS009D0
c 19	75	2.1	1201	9	AL069706 Drosophila
c 20	74.8	2.1	1268	28	BZ577630
c 21	74.6	2.1	1195	13	AL069706 Drosophila
c 22	74.4	2.1	1101	29	CNS0039G
c 23	74.2	2.1	576	29	CNS035N7
c 24	73.8	2.1	1098	13	AL069706 Drosophila
c 25	73.8	2.1	1101	29	CNS01219
c 26	73.6	2.1	1201	13	AL069706 Drosophila
c 27	73.6	2.1	1201	13	AL069706 Drosophila
c 28	73.4	2.0	996	29	CNS009D0
c 29	73.2	2.0	870	29	CNS009D0
c 30	73.2	2.0	1175	28	BZ696793
c 31	73.2	2.0	1201	13	AL069706 Drosophila
c 32	73	2.0	1201	13	AL069706 Drosophila
c 33	72.8	2.0	991	29	CNS0012D
c 34	72.8	2.0	997	13	AL069706 Drosophila
c 35	72.8	2.0	1084	29	CNS06FLB
c 36	72.8	2.0	1200	13	AL069706 Drosophila
c 37	72.8	2.0	1200	29	CNS016EL
c 38	72.6	2.0	614	29	CNS0152H
c 39	72.6	2.0	881	29	CG767277
c 40	72.6	2.0	1201	13	AL069706 Drosophila
c 41	72.6	2.0	1348	29	CG749499
c 42	72.4	2.0	706	29	CE784276
c 43	72.4	2.0	1061	13	AL069706 Drosophila
c 44	72.4	2.0	1139	28	AQ897537
c 45	72.4	2.0	1165	13	AL069706 Drosophila

## ALIGNMENTS

RESULT 1  
CNS00EVL/c

LOCUS  
DEFINITION

CNS00EVL 1101 bp DNA linear GSS 04-JUN-1999  
Drosophila melanogaster genome survey sequence T7 end of BAC.  
BAC29B23 of RFL-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
- Web : www.genoscope.cns.fr

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosier in Pieter de Jong's laboratory in the Department of



```

RESULT 3
CNS012FZ/c
LOCUS
DEFINITION
  CNS012FZ 1094 bp DNA linear GSS 26-JUL-1999
  Drosophila melanogaster genome survey sequence T7 end of BAC
  BACN07B02 of DrosBAC library from Drosophila melanogaster (fruit
  fly), genomic survey sequence.
ACCESSION
  AL101513.1 GI:5613124
VERSION
  AL101513.1
KEYWORDS
  Drosophila melanogaster (fruit fly)
ORGANISM
  Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
  1 (bases 1 to 1094)
REFERENCE
  Genoscope.
  Direct Submission
  Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
  BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
  - Web : www.genoscope.cns.fr)
  Determination of this BAC-end sequence was carried out as part of a
  collaboration with the European Drosophila Genome Project (EDGP) -
  http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
  library (Dros BAC) was made by Alain Billaud at CEPH (Centre
  d'Etude du Polymorphisme Humain) with funding provided by a MRC
  project grant. The DNA was prepared from embryos by Alain Bucheton
  and Genevieve Payan. It has been constructed in the vector
  pBelOBAC11.
FEATURES
  Location/Qualifiers
  1..1094
   /organism="Drosophila melanogaster"
   /mol_type="genomic DNA"
   /db_xref="taxon:7227"
   /clone="BACN07B02"
   /clone_lib="DrosBAC"
   /plasmid="pBelOBAC11"
   /note="end : T7"

ORIGIN
  Query Match 2.3%; Score 82; DB 29; Length 1094;
  Best Local Similarity 36.4%; Pred. No. 1.8e-07;
  Matches 192; Conservative 88; Mismatches 246; Indels 2; Gaps 1;

Qy 558 TATTAAATAACAGCTTCTAAAGCGTCTTATTGGATGAATTCGAAATATCAATAA 617
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1090 WAAAAAAGATGGAARANKGTWDTTTTIDWMTTGGAAAAAGAAAAATTTTWT 1031

Qy 618 TAATTGATGCTATTATTACTTGCTGTTATGGTATCACTTTCATGCTCTATACATGTAAT 677
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1030 TGRWTTTAKRWTTTITKAWTTTTTTTADWRRAAATTTTIDRWTAATATWATWTTT 971

Qy 678 ATATTTCGAGTAGACCTTAATCAAGTAAATTTGCTATTATTATTTATTCGAATAAT 737
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
970 TTWTWTGAGRGARAAATWAAATWAAAAAWWWATWAAATWAAADAGRTKAWTTTAG 911

Qy 738 ATGTAATCGATGCTTTGTGG--TTATTTTATGCTTTGTTTCATTTTAAATGACGGTGAG 795
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
910 GRTTGRGTGRTTWTTAARGRTWAAAWTTTWTGATTTTWTWTTGAAAGAWTTTAWW 851

Qy 796 CTTCGCAATTCATATTTTATGATGACAACTCTTTGATGAAGTATTTAAGATATTTGTT 855
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
850 TTTTWTWTTTTAAATTTWATTTAAWTAATTTTATTAARWAWWTTTITAWADTTTT 791

Qy 856 AATGATGAGGGTTGCGGTATTTTATATTAATATCAATATAAATCAACAAATATAT 915
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
790 TTTTTRRRRTTGRWTTTAAATTTAAAGAGTTTWTWTTTITWAAAAAATATATTT 731

Qy 916 GTTATTTTGTCTTTTATAGTGTCTTTTAAAGAGGTAGGATGACCTAAAGTTCGCT 975
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
730 TTTTWTWTTWWTTTITTTTAAAAAATAAARTTTTIDGGGTGRTTWTWATATAATT 671

Qy 976 AAATATGCGTAATGTCATTGCTATTAATTCACCTCAAGATACACTATTGGCAATG 1035

```

```

Db 670 WAAKTTTGGRWARTTTTTTTRRAAAWWTGAWATGAGVTTTGRITGGRKAAATKG 611
Qy 1036 ACAATAATGTCACCTCGTATGAACAATATTAGTAGATGTTGTTTTC 1083
Db 610 TWAATTTTGGTTGGAAATTAAGRRWTTTATTTTAAATTTTTTTT 563

RESULT 4
EX437739
LOCUS
DEFINITION
  EX437739 Homo sapiens THYMUS Homo sapiens cDNA clone CSOCAP008YG24
  3-PRIME, mRNA sequence.
ACCESSION
  EX437739
VERSION
  EX437739.1 GI:31018315
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
  1 (bases 1 to 1200)
REFERENCE
  Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
  Full-length cDNA libraries and normalization
  Unpublished (2001)
  Contact: Genoscope
  Genoscope - Centre National de Sequencage
  BP 191 91006 EVRY cedex - France
  Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
  Library was constructed by Life Technologies, a division of
  Invitrogen. This sequence belongs to sequence cluster 534.r For
  more information about this cluster, see
  http://www.genoscope.cns.fr/
  cgi-bin/cluster.cgi?seq=CSOCAP008BD12NP1&cluster=534.r. Contact :
  Feng Liang Email : fliang@lifetech.com URL :
  http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
  Faraday Avenue Genoscope sequence ID : CSOCAP008BD12NP1.
FEATURES
  Location/Qualifiers
  1..1200
   /organism="Homo sapiens"
   /mol_type="mRNA"
   /db_xref="taxon:9606"
   /clone="CSOCAP008YG24"
   /tissue_type="THYMUS"
   /clone_lib="Homo sapiens THYMUS"
   /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
   with a NotI-oligo(dT) primer. Five prime end enriched,
   double-strand cDNA was digested with Not I and cloned into
   the Not I and EcoRV sites of the pCMVSPORT 6 vector.
   Library was not normalized."

ORIGIN
  Query Match 2.2%; Score 80.2; DB 13; Length 1200;
  Best Local Similarity 35.3%; Pred. No. 4.5e-07;
  Matches 270; Conservative 112; Mismatches 379; Indels 4; Gaps 2;

Qy 515 TTGGTGGCTTAAATCGTTACTGCTCATATTCATTAATTCATTAATTAACAGTTC 574
Db 416 TTATTTTATTTATTAATTAATTTTATTAATTTATTTATTTATTTATTTATTTT 475

Qy 575 TAAAGCGTGTATTGGATGAATTCGAAATTAATCAATATTAATTCATGCTATTAT 634
Db 476 TTATTAATTTTATTTTAAATTAATTTTATTTTATTTTATTTTATTTTATTTT 535

Qy 635 ACTTGCTGATTGGTATCAACTTTTCATGCTCTATACATGTAATATTTTCGAGTTAGACC 694
Db 536 TATTATKTTTAAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTT 595

Qy 695 TTAATTAAGGTAATTTGCTATTATTAATTAATTAATTAATTAATTAATTAATTAAT 754
Db 596 ATTTTTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 655

Qy 755 GTGGTATTTTATGTTTGTTCATTTTAAAGACGCTGAGCTTGTGCAATTCATATTTT 814

```



```

/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="nbx50003600r"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/notes="Vector: pBelOBAC11, Site 1: HindIII; Site 2:
HindIII; Rice is one of two most popular grains in the
world. Half of the world population especially those
inhabiting highly populated areas of the humid tropics
and subtropics, rely on rice as their primary source of
carbohydrate. Monocotyledonous rice is a diploid plant
(2n=24) with a haploid genome equivalent of 431 Mbp
(Arumuganathan and Earle, 1991). The relatively small
genome of rice, three times larger than that of
Arabidopsis, makes it suitable for genomic studies. In
order to facilitate positional cloning, physical mapping
and genome sequencing of rice, we have constructed a BAC
library from Oryza sativa, Nipponbare variety. The
library contains 36,864 clones with an average insert size
of 128.5 Kb providing 10.9 haploid genome equivalents. The
deep coverage allows the isolation a particular sequence
with a probability of 99.9 %. Two high density filters,
each containing 18,432 clones (doubly spotted), represent
the whole library for colony screening."

```

## ORIGIN

```

Query Match      2.2%; Score 78.8; DB 28; Length 714;
Best Local Similarity 48.9%; Pred. No. 9.4e-07;
Matches 209; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

```

```

QY 519 GTGCTTTAATCGTACTGCTCATTTCAATTAATTCACATATTAATTAACAGTTCTAA 578
Db 195 GTATATNAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 254
QY 579 AGGCTGTTTATCGATGAATTCGAAATTCACATAATTAATGATGCTTAATTAATTT 638
Db 255 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 314
QY 639 GCTGATGTTGATCAATTTTCATGCTATPACATGAATTAATTTTCAGTTAGACCTTAA 698
Db 315 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 374
QY 699 TTCAGGTAATTTGCTATTTTATTTATTTATTCGAATTAATGTAATCGATGCTTTGCG 758
Db 375 ATTAATTTTATTTATTTATTTTATTTATTTATTTTATTTTATTTTATTTTATTTA 434
QY 759 TTTATTTTATGTTTGTTCATTTTATGACGGTGAGCTTGTCATTTTCATTTTATG 818
Db 435 TTATATTTTATTTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 494
QY 819 ATGACAACTCTTTGATGAAGTATTTAAGATATTTGTAATGATGAGGGTTTGGGTGA 878
Db 495 TTTTATATTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 554
QY 879 TTTTATATTTAAATCATATAAAATCAACAATATATGTTATTTTGTGCTTTTATAGT 938
Db 555 TTTTATTTTATTTTATTTTATTTATTTATTTATTTATTTTATTTTATTTTATTTT 614
QY 939 GTTCTTT 945
Db 615 ATTTTTT 621

```

```

RESULT 7
AL536104 1201 bp mRNA linear EST 31-MAY-2003
LOCUS
DEFINITION AL536104 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
CSODF022YC18 5-PRIME, mRNA sequence.
ACCESSION AL536104
VERSION AL536104.2 GI:31260974
KEYWORDS EST.

```

```

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li,W.B., Gruber,C., Jesses,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On Feb 13, 2001 this sequence version replaced gi:12799597.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: sefre@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CSODF022B89QP1.

```

## FEATURES

source

```

1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODF022YC18"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/notes="Organ: brain; vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo (df) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

```

## ORIGIN

```

Query Match      2.2%; Score 77.4; DB 9; Length 1201;
Best Local Similarity 34.4%; Pred. No. 1.9e-06;
Matches 153; Conservative 93; Mismatches 197; Indels 2; Gaps 1;

```

```

QY 523 TTTATTCGTTACTGCTCATATTCAAATTAATTCATTAATAAACAGTTCTAAAGCC 582
Db 745 TAWTTTWTATATATTTTAAATTTTATATTTTATRWATATTTTWTWTWTWTAT 804
QY 583 TGTTTATTTGGATGAATATTCGAAATTCACATATTAATTTGATGCTATTTACTTGCTG 642
Db 805 TAAATWATAATWTTAAWTTTWTATTTTATWATAAAWTTT-TWATATTTWTATTTTWA 862
QY 643 TATTCGTTATCACTTTTCATGCTCATACATGTAATATATTTTCGAGTTAGACCTTAATCA 702
Db 863 TAAATTTTATTTTAAATTTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTT 922
QY 703 AGGTAATTTGTTCTATTTTAAATTTATTTATCTGAATAATATGATTCGATTCGTTGTTAT 762
Db 923 WWWWAAATWATTAGTAAWAAWAAATATATATATATATATATATATATATATATATAT 982
QY 763 TTTTATGTTTGTTCATTTTAAACACGGTGAGCTTGTCATTCATATTTTATGATGA 822
Db 983 WTAAAAAAWAAATWAAATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1042
QY 823 CAACATCTTTGATGAAGTATTTAAGATATTTGATGATGAGGGGTTTGGGTTATTTT 882
Db 1043 TATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1102
QY 883 TTTATTTAAATCAATTAATAAATCAACAATATATGTTATTTTGTGCTTTTATAGTTC 942
Db 1103 AWWAAAAAATAAATAADAWATATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1162

```

```

RESULT 8
BX416727
LOCUS

```

```

QY 943 TTTTAAAGAGGTAGGATGACCTTAA 967
Db 1163 TATTAAWWTAGTKRTATATATAWA 1187

```

```

BX416727 712 bp mRNA linear EST 15-MAY-2003

```

```

DEFINITION BX416727 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
ACCESSION CS0DA011Y114 5-PRIME, mRNA sequence.
VERSION BX416727
KEYWORDS BX416727.1 GI:30765629
SOURCE EST. Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 712)
JOURNAL Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
COMMENT Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DA011B07QP1.
FEATURES
Location/Qualifiers
source
1..712
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DA011Y114"
/tissue_type="NEUROBLASTOMA"
/clone_lib="Homo sapiens NEUROBLASTOMA"
/notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
ORIGIN
Query Match 2.2%; Score 77.2; DB 13; Length 712;
Best Local Similarity 21.7%; Pred. No. 2.1e-06;
Matches 91; Conservative 153; Mismatches 176; Indels 0; Gaps 0;
QY 541 ATATTCATTAATTCATTAATAAACAAGCTCTAAAGGCTGTTTATGATGAATAT 600
DB 240 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 299
QY 601 TCGAATATACATAATATGATGCTATTAATCTGCTGATGCTGATGCTCACTTCA 660
DB 300 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 359
QY 661 TGCTCTACATGATATATTCGAGTTAGACCTTAATCAAGGTAATTTGCTATTTA 720
DB 360 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 419
QY 721 ATTAATATCAATAATATGATGCTGTTGCTGTTGTTGTTGTTGTTGTTGTTGTT 780
DB 420 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 479
QY 781 TTTAATGAGGTGAGCTGTGCTATTTATTTTATGATGATGATGATGATGATGATG 840
DB 480 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 539
QY 841 ATTTAAGATATTTGATGATGAGGTTGCTGTTGCTGTTGCTGTTGCTGTTGCT 900
DB 540 KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 599
QY 901 AAATCAACAATATATGTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 960
DB 600 KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 659
RESULT 9
CNS0039G/c 1101 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TBT3 end of BAC #
DEFINITION

```

```

BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
REFERENCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
AUTHORS Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
TITLE Ephydroidea; Drosophilidae; Drosophila.
COMMENT 1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammos in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
Location/Qualifiers
source
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08K10"
/clone_lib="RPCI-98"
/notes="end : TBT3"
ORIGIN
Query Match 2.1%; Score 76.6; DB 29; Length 1101;
Best Local Similarity 17.5%; Pred. No. 2.8e-06;
Matches 124; Conservative 299; Mismatches 282; Indels 2; Gaps 1;
QY 614 ATAATAATGATGCTATTACTTGTGCTGATGCTGATCACTTCTCATGCTTACATG 673
DB 1091 WDRTRKDDWMTKWTWKDRADRRWAGDADRWADDDGAGTWTATWWWWWWATWD 1032
QY 674 TAATATATTCGAGTTAGACCTTAATTCAGGTAATTTGCTATTTATTTATTTATCTGAA 733
DB 1031 TWWDKWWWWATAAKTDTANTWRTAWRADWAGRGAGKGRDRDAATDADGAGRGGRKR 972
QY 734 TAATATGTAATGATGCTTTGCTGTTATTTTATGTTTGTTCATTTTATGACGGTG 793
DB 971 KKKRDKDDDDKKGGKKKAAKAAKATKWDWDWDKWDKWDKADKDDDDGADGDK 912
QY 794 ACCTTGCTCATTTATTTTATGATGACACATCTTTGATGAAGTATTTAAGATATTG 853
DB 911 DDDGKGDADDTDGTDKDDDKWDWDWDKAGTWDGATWAAATDWWGWDADWWTW 852
QY 854 TTAATGCAATGAGGGTTGCGTGATT--TTTATTAATTAATCAATAATAATCAACAAT 911
DB 851 DAAADDWWADRDWAWAKWDDAWAGARTADRDWGDGKRGKGRKRRKDDKDDK 792
QY 912 ATATGCTATTTTGTCTTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTT 971
DB 791 AADDDDAATTTTWTTRTDDKWKTDWTWADRTWDRDDDDDDDRBAGTAGRKR 732
QY 972 GCCTAAATATGGCTAAATGCGCATTTGCTATATATCACTCAAGATACACTATTGGCA 1031
DB 731 RTWKRRWRDRTRDDADADTDARDDRRRRGGDADAGKGGKTKGRKRRDRATWDR 672

```

```
QY 1032 ATTGACAAATATGTCACCTTCGTATGAACAATATTAGTAGATGTTGTTTTCCTGCAAAA 1091
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 671 WADAAGAAATTTTDTDDDDKRRRRGARRRRRTTARAADWWTWKADWAKWDKTRAD 612
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1092 ATAAAAAATTTTCTGTTGAAATAAATCAAGCCCTCTAGCGTTTTCCTTATCTTAAAT 1151
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 611 RWDRAADTWDARKADRDWAKARAWARRDRARAARDRRRTTKGKTATATTTTAAARA 552
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1152 ACAGGAATAGCAATGAAGTTAAATTGACACTTAAGCAATAGTCAACCTTAACAGGAG 1211
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 551 AWMAMWAAATTTATTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT 492
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1212 GAACCTATGCTTGTCAAGAGCATCAATTTGACCAACTTCTTAAACCTCTGAGTGAT 1271
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 491 AAAAAAATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 432
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1272 TCGATCTGGCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1318
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 431 WAATTTATTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT 385

RESULT 10
CNS00HU4 1101 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC:
DEFINITION BACR35018 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL074046.1 GI:4953725
VERSION 1
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
REFERENCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (02-JUN-1999): Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
Location/Qualifiers
1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR35018"
/clone_lib="RPCI-98"
/note="end : 17"

FEATURES
source
Query Match 2.1%; Score 76.6; DB 29; Length 1101;
Best Local Similarity 34.7%; Pred.No.2.8e-06;
Matches 191; Conservative 11; Mismatches 348; Indels 0; Gaps 0;

QY 628 TATATTAATCTGCTGATGATGATCACTTTCATCTCTATACATGAATATTTTCGAG 687
Db 48 TTTTCTTTTCTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 107
```

```
QY 688 TTAGACCTTAATCAAGGTAAATTTGCTCTATTTATTTATTTATCTGAATAATGTAATCGA 747
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 108 TTTGCTDGDGATGATNAATAAAAAAATTTAAWATTTATTTATATTTCHDWTNTTTTATTIN 167
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 748 TTGCTTTGCTGATTTATTTTATTTGTTGTTTCTTCTTCTTCTTCTTCTTCTTCTTCT 807
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 168 TTTTNTNTNTTATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 227
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 808 TATTTTATGATGACAAACATCTTTGATGAAGTATTTAAAGATATTTGTTAATGCAATGAGG 867
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 228 TTTTNTNTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 287
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 868 GTTTCGCGTATTTTATTTATTAATCATATAATAAAATCAACAATATATGTTATTTGCT 927
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 288 NNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNT 347
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 928 CTTTNTTATAGTCTTTTAAAGAGGTAGGATGACCTAAAGGTGCGCTTAAATATGCGCTA 987
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 348 TTTTATTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 407
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 988 AATTGCCATCTGCTATAATTCACCTCAAGATACACTATTTGGCAAAATTCGACAAATATGTC 1047
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 408 NNAATTTAATTTTNTTATMAAATANATTTTAAATTTATGTTNNTNNTNNTNNTNNTNNT 467
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1048 CTTGCTATGAACAATATTTAGTAGTGTGTTTCTGCTGCAAAATATAAAATTTTCTGG 1107
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 468 NTATANTNNTNNTTACANNANNAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTNN 527
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1108 TTGAATAAATCAAGGCTCTAGCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1167
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 528 TTTAANNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNT 587
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1168 GAAGTTAAT 1177
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 588 TTTNTTNTT 597
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
BX439779/1 1201 bp mRNA linear EST 15-MAY-2003
LOCUS BX439779 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE014YF05
DEFINITION 3-PRIME, mRNA sequence.
ACCESSION BX439779.1 GI:30771778
VERSION 1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1201)
Li W.B., Gruber C., Jessee J. and Polayes D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 3370.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DE014CC03NP1&cluster=3370.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paraday Avenue Genoscope sequence ID : CS0DE014CC03NP1.
Location/Qualifiers
1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DE014YF05"
/tissue_type="PLACENTA"
```





```
source
1. 1200
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0CAP008Y104"
/tissue_type="THYMUS"
/clone_lib="Homo sapiens THYMUS"
/note="Vector: pCMVSPORT6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."

ORIGIN
Query Match 2.1%; Score 76.2; DB 13; Length 1200;
Best Local Similarity 36.0%; Pred. No. 3.5e-06;
Matches 247; Conservative 86; Mismatches 352; Indels 2; Gaps 1;

QY 523 TTTAAATTCGTTACTGCTCATATTCATTAATTCACATTAATTAATAAAGAGTCTCTAAAGGC 582
Db 497 TTAATTTTAAAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 556
QY 583 TGTATTGATGAATATTCGAATATTCACATTAATTAATTAATTAATTAATTAATTAATTAAT 642
Db 557 AATATTTTAAAAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 616
QY 643 TATTGGTATCACTTCATGCTCTATACATGTAATATATTCGAGTTAGACCTTAATTC 702
Db 617 TATTTTATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 676
QY 703 AGGTAATTTGCTATTTAATTTATCTGAATATTAATTAATTAATTAATTAATTAATTAATTT 762
Db 677 TTTTAAATTAATTAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 736
QY 763 TTTTATGTTTGTTCATTTTAAAGCGGTGAGCTGTCGATTCATATTTTATGATGA 822
Db 737 AWTATTAATACATANATATATATATATATATATATATATATATATATATATATATATTT 796
QY 823 CAACATCTTTGATGAAGTATTTAAGATATTTAATGTCATGAGGGTTTCGCGTATTTT 882
Db 797 TWACAAAAATWAAAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 856
QY 883 TTATTTAAATCATATAAATAATCAATATATATTTTGTGCTCTTTTATAGTGTTC 942
Db 857 TTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 916
QY 943 TTTTAAAGAGGTAGATGACCTAAAGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTAT 1002
Db 917 TTAAATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 976
QY 1003 AATTCACCTCAAGATACACTATTGGCAATTTGCAAAATATGTCATCTGCTATGAACAA 1062
Db 977 ATANNTAATWAAATWAAAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 1036
QY 1063 TATTAGTAGATGTTGTTTGTGCTGCAAAATTAATTTTTCGCTGCTGCAATACTCA 1120
Db 1037 TATTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 1096
QY 1121 AGGCTCTAGGTTTCTCTTATCTTAAATACAGGAATAGGAAATAGGAAATAGGAAATAGG 1180
Db 1097 TTATTTAAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 1156
QY 1181 ACTTAAGCAATAGTCAACCTAACAGA 1207
Db 1157 WTAAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 1183

RESULT 14
BX437739/c 1200 bp mRNA linear EST 22-MAY-2003
LOCUS BX437739 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP008Y24
DEFINITION 3-PRIME, mRNA sequence.
ACCESSION BX437739
```

```
VERSION BX437739.1 GI:31018315
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
MAMMALIA; EUTHERIA; PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
REFERENCE 1 (bases 1 to 1200)
AUTHORS Li W.B., Gruber C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 534.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0CAP008BD12NP1&cluster=534.r. Contact :
Feng liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0CAP008BD12NP1.
FEATURES
Location/Qualifiers
1..1200
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0CAP008Y24"
/tissue_type="THYMUS"
/clone_lib="Homo sapiens THYMUS"
/note="Vector: pCMVSPORT6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."

ORIGIN
Query Match 2.1%; Score 75.8; DB 13; Length 1200;
Best Local Similarity 36.9%; Pred. No. 4.2e-06;
Matches 246; Conservative 93; Mismatches 322; Indels 5; Gaps 2;

QY 542 TATTCAATTAATTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 601
Db 1193 TTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 1134
QY 602 CGAAATTAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 661
Db 1133 TTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 1074
QY 662 GCTCTATACATGATATATATTCGAGTTAGACCTTAATTAATTAATTAATTAATTAATTAAT 721
Db 1073 TTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 1014
QY 722 TTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 781
Db 1013 TTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 954
QY 782 TTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 837
Db 953 TTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 894
QY 838 AGTATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 897
Db 893 AATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 834
QY 898 ATAAATCAACAAATATATTTTGTCTCTTTTATAGTCTCTTTTAAAGAGGTAGG 957
Db 833 ATTAAATCAACAAATATATTTTGTCTCTTTTATAGTCTCTTTTAAAGAGGTAGG 774
QY 958 ATGACCTAAAGTCGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1017
Db 773 ATAAATCAACAAATATATTTTGTCTCTTTTATAGTCTCTTTTAAAGAGGTAGG 715
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2004, 22:56:15 ; Search time 77 seconds  
(without alignments)  
1698.954 Million cell updates/sec

Title: US-10-780-347-2

Perfect score: 2343

Sequence: 1 MFLSKHQIQLSKPLSDSI.....DTQSKQDKPQSSATSALSW 463

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

- 1: Geneseqp1980s:\*
- 2: Geneseqp1990s:\*
- 3: Geneseqp2000s:\*
- 4: Geneseqp2001s:\*
- 5: Geneseqp2002s:\*
- 6: Geneseqp2003as:\*
- 7: Geneseqp2003bs:\*
- 8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	264.5	11.3	364	6	ADA35354 Acinetoba
2	211.5	9.0	376	6	Abm67333 Photorhab
3	152.5	6.5	1454	7	Add93813 Human apo
4	152.5	6.5	1503	2	Aaw48845 Human rec
5	152.5	6.5	1503	7	Add93815 Human HJO
6	143	6.1	884	6	Aao19960 C elegans
7	138	5.9	2008	2	Aaw22016 Utrrophin
8	136.5	5.8	7201	4	Abb71136 Drosophil
9	135	5.8	2013	4	Abb71136 Amino aci
10	134.5	5.7	1179	4	Aau37669 Streptoco
11	134.5	5.7	1179	4	Aam01107 CFE llo p
12	134.5	5.7	1179	6	Abu46083 Protein e
13	131	5.6	1329	7	Adel15650 Human str
14	129.5	5.5	860	7	Adc38517 Human AML
15	129	5.5	1328	6	Abp14658 Novel hum
16	128	5.5	2017	4	Abg08301 Novel hum
17	127.5	5.4	624	6	Abm66947 Photorhab
18	127	5.4	1294	6	Abu33368 Protein e
19	127	5.4	1726	6	Aao16416 Human nuc
20	127	5.4	2278	6	Abu39818 Human SCA
21	126.5	5.4	752	5	Aau84329 Protein D
22	126.5	5.4	752	6	Abu92070 Human cer
23	126.5	5.4	785	6	Abu92070 Human cer
24	126.5	5.4	893	6	Abu92070 Human cer
25	126.5	5.4	1133	6	Abu92070 Human cer

26	126.5	5.4	1133	6	ABR62252	Abp62252 GPBP-inte
27	126.5	5.4	1135	6	ABR62251	Abp62251 GPBP-inte
28	126	5.4	610	2	AAW68206	Aaw68206 M. catarr
29	124.5	5.3	1940	5	ABG79661	Abg79661 Invertebr
30	124.5	5.3	1940	6	ABU89712	Abu89712 Protein d
31	123	5.2	1935	7	ABU89712	Abu89712 Protein d
32	122	5.2	1370	4	ABD45207	Abd45207 Rat Prote
33	122	5.2	1370	4	ABD45207	Abd45207 Rat Prote
34	121.5	5.2	1179	6	ABG07508	Abg07508 Novel hum
35	121.5	5.2	1179	6	ABU01677	Abu01677 S. pneumo
36	121.5	5.2	3259	7	ABU01677	Abu01677 S. pneumo
37	121	5.2	3259	7	ABU01677	Abu01677 S. pneumo
38	120.5	5.1	1048	4	ABU20426	Abu20426 Human pro
39	120.5	5.1	1294	4	ABU20426	Abu20426 Human pro
40	120	5.1	1131	4	ABG07281	Abg07281 Drosophil
41	120	5.1	1497	4	ABG20153	Abg20153 Novel hum
42	120	5.1	1935	5	ABG31649	Abg31649 Amino aci
43	119.5	5.1	1066	4	AAG67418	Aag67418 Amino aci
44	119.5	5.1	1066	6	ABG72693	Abg72693 Fruitfly
45	119.5	5.1	1179	4	AAU38015	Aau38015 Streptoco

ALIGNMENTS

RESULT 1  
ADA35354  
ID ADA35354 standard; protein; 364 AA.  
XX  
AC ADA35354;  
XX  
DT 20-NOV-2003 (first entry)  
XX  
DE Acinetobacter baumannii protein #2515.  
XX  
KW Acinetobacter baumannii; bacterial disease; vaccine;  
KW plant biocontrol agent.  
XX  
OS Acinetobacter baumannii.  
XX  
PN US6562958-B1.  
XX  
PD 13-MAY-2003.  
XX  
PF 04-JUN-1999; 99US-00328352.  
XX  
PR 09-JUN-1998; 98US-0089701P.  
XX  
(GENO-) GENOME THERAPEUTICS CORP.  
XX  
PA Breton G, Bush D;  
XX  
PI WPI; 2003-576092/54.  
XX  
DR N-PSDB; ADA31228.  
XX  
PT New Acinetobacter baumannii proteins and nucleic acids, useful as reagents  
PT for diagnosing a bacterial disease, as components of antibacterial  
PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for  
PT plants.  
XX  
PS Example; SEQ ID NO 6641; 328pp; English.  
XX  
CC The invention relates to isolated Acinetobacter baumannii nucleic acids.  
CC The A. baumannii nucleic acids and polypeptides are useful as reagents  
CC for diagnosing a bacterial disease, as components of antibacterial  
CC vaccines, as targets for antibacterial drugs, to detect the presence of  
CC A. baumannii and other Acinetobacter species in a sample, in screening  
CC compounds for the ability to interfere with the A. baumannii life cycle  
CC or to inhibit A. baumannii infection, and as biocontrol agents for  
CC plants. The present sequence represents the amino acid sequence of an A.  
CC baumannii protein.  
XX  
SQ Sequence 364 AA;

Query Match 11.3%; Score 264.5; DB 6; Length 364;  
 Best Local Similarity 21.5%; Pred. No. 2.3e-13;  
 Matches 87; Conservative 76; Mismatches 152; Indels 89; Gaps 14;

QY 8 IEQLSKPLSDSDSCGVYLKLEKSAFRPLNEFVAQTALRKLSONPSADERDALOEACLN 67  
 DB 5 ISELKPLINDSLCC-----EDYSF---SNEFHEIKKA--RTQDDLLDQGDWVAERKQA 54  
 QY 68 KWKILSDSLYQFSKTTREDIELISWFAAQFLDITLESAAANSLEWLAJLSEKHWDHLP 127  
 DB 55 DNDPFAKSVSTLLIEKTKDIELLWVIAWTHLN--GEGGVKGTILTHMLNQWQDIHP 113  
 QY 128 VLPVETLKSDDKGKEREQADAKVAFQVLVDSESSILYAPVL--OLPLNGEVTPFF-- 183  
 DB 114 II-----EDDD-----LDQRIGLLQGLNQ-----LPMLLKKVPLTNTAPYNL 152  
 QY 184 -----DFOAERKGRISOLKSMILTTVAQERFAIQFVNAKRCVTC 225  
 DB 153 LDYDNFLYHENIRKQTEYESQSGPSELEQDAIFNT-----SKTFQSVNYQBFNSV 206  
 QY 226 LDRLSALYSTKCHSLGSGSTNFGFAKSLLTREVENAL-----VHLSGIKLAPKAEAKTVEQ 280  
 DB 207 LTEWNLKQTLDELHGLDPSFAAIDGAFETIHSTLRKIVKAEAFGTGLAPSOEQAAV-- 264  
 QY 281 EVASSVSEGEPLPSHMDTKHTERIPMAEQATVSQHLHAGNLSELGNLNNMRDLAFHL 340  
 DB 265 -----ITPSMENQVQVIVSDQFMFPQQAQT---HL-----ANREQAMKV 302  
 QY 341 LREVSDFRQSEPHSPISFLLEKAIKRWGLSLPPELLREMMSEQN 384  
 DB 303 LQEIADYFQANEHSPSVYMLQTKIKSQMPLHWAQVIAKDEH 346

RESULT 2  
 ASM67333  
 ID ASM67333 standard; protein; 376 AA.  
 XX  
 AC ASM67333;  
 XX  
 DT 20-NOV-2003 (first entry)  
 XX  
 DE Photorhabdus luminescens protein sequence #430.  
 XX  
 KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
 KW detection; food; gene expression; plant; animal; microorganism; toxin;  
 KW antibiotic; biopesticide; virulence factor; disease model; plague;  
 KW whooping cough.  
 XX  
 OS Photorhabdus luminescens.  
 XX  
 PN WO200294867-A2.  
 XX  
 PD 28-NOV-2002.  
 XX  
 PF 07-FEB-2002; 2002WO-IB003040.  
 XX  
 PR 07-FEB-2001; 2001FR-00001659.  
 XX  
 PA (INSP ) INST PASTEUR.  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;  
 PI Buchrieser C;  
 XX  
 DR WPI; 2003-148459/14.  
 XX  
 PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
 PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.  
 XX  
 PS Claim 2; SEQ ID NO 430; 1205pp; French.  
 XX  
 CC The invention relates to the isolation of genes and their encoded

proteins from Photorhabdus luminescens. The isolated sequences are  
 sources of probes and primers for detecting the genome of P. luminescens  
 and related species; to study polymorphisms; for gene analysis and for  
 detection/amplification of the genes. Antibodies (Ab) raised against the  
 polypeptides encoded by the genes are used for detection/identification  
 of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that  
 carry a gene-containing vector are used to select compounds that  
 modulate, regulate, induce or inhibit expression of the genes in plants,  
 animals or microorganisms other than P. luminescens and are able to alter  
 response or sensitivity to toxins and antibiotics produced by P.  
 luminescens. Cells transformed to express the genes are useful for  
 recombinant production of the proteins, particularly toxins and  
 antibacterials useful as insecticides, bactericides and fungicides. The  
 genes, proteins, vectors containing the genes and Ab are also useful  
 therapeutically (to treat microbial infection by bacteria or fungi that  
 are sensitive to P. luminescens-encoded toxins or antibiotics) and as  
 biopesticides. Other uses of the genes and the proteins are as virulence  
 factors and for identifying targets of human diseases for which P.  
 luminescens is a model (particularly plague and whooping cough). This  
 sequence represents one of the isolated P. luminescens proteins

XX  
 SQ Sequence 376 AA;

Query Match 9.0%; Score 211.5; DB 6; Length 376;  
 Best Local Similarity 22.0%; Pred. No. 6.5e-09;  
 Matches 96; Conservative 75; Mismatches 159; Indels 107; Gaps 17;

QY 3 LSKHQIEQLSKPLSDSDSCGVYLKLEKSAFRPLNEFVAQTALRKLSONPSADERDALQ 62  
 DB 11 LAQTDIARFLQPIAADKPAAGCDIEYE-----PIFEQINAARET-----DDDFLQ 54  
 QY 63 -----EACLNKWKILSDSLYEQFSKTRDIELISWFAAQFLDITLESAAANSLEWLAJ 116  
 DB 55 DDTWGYETRQADWKMVYALCOEVLKQSKDQIACWLTQAQOGL-YGLAGISGGITLLAH 113  
 QY 117 LSEKHWDHLPVLPVETLKSDDKGKEREQADAKVAF-----QLVGDSESSILYAPVL 172  
 DB 114 LLETTPVLYPPL-----DD--TEANSADARLGRSLWLDNQVLKQDNLTLTDDGKL 163  
 QY 173 QLPLAGEVTFDFQSAERKGRISOLKSMIL-----TTTVAQERFAIQPOME 217  
 DB 164 SLUSVQRVOYFE-----QRAAVNSELSALISDGVFGMAECDGSIPTTFAEPNQLLRAE 219  
 QY 218 NAKRCVTQDLRLSA-LVSTKCHSLGSGSTNFGFAKSLLTREVENALVHLSGIKLAPKAEAK 276  
 DB 220 QVNKALAEQLKIMVELLPDAGNSMSASTORLQELVALIERFRD-----WVAFGS--- 268  
 QY 277 TVEQEVASS-----VSEGEPLPSHMDTKHTERIPMAEQATVSQHLHAGNLSELGNLNNMN 333  
 DB 269 --SQEYGSNEAIAAGGAPNGMTTFL-----TEDRAHHEVRAIGOMIN-- 313  
 QY 334 RDLAFHLLREVSDFRQSEPHSPISFLLEKAIKRWGLSLPPELLREMMSEQNGDALSTIFN 393  
 DB 314 -----IANYFRQNEPTSPVYLIERAARWANNMCMGMAEWLEWKE-----N 353

QY 394 AAGLNHLDOVLL-PEVS 409  
 DB 354 TSAEQEIMRMVKGPEIN 370

RESULT 3  
 ADD93813  
 ID ADD93813 standard; protein; 1454 AA.  
 XX  
 AC ADD93813;  
 XX  
 DT 29-JAN-2004 (first entry)  
 XX  
 DE Human apoptosis/nerve growth-related protein HJ06972.  
 XX  
 KW Human; HJ06972; KIAA1079; AATYK homologue;  
 KW apoptosis-associated tyrosine kinase; apoptosis;  
 KW nerve cell differentiation; nerve cell proliferation; nerve cell growth;

KW nerve cell division; drug screening; transgenic animal;  
KW apoptosis disorder; nerve regeneration; neurological disorder; hepatitis;  
KW hepatotropic; anti-inflammatory; virucide; cytostatic.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Region 1..1443  
FT /note= "This region is common to HJ06972 and its splice  
FT variant AF06972"  
FT 1444..1454  
FT Region /label= HJ06972\_specific\_C-terminal\_sequence

XX W02003080836-A1.

XX 02-OCT-2003.

XX 26-MAR-2003; 2003WO-JP003713.

XX 26-MAR-2002; 2002JP-00086843.

XX (KAZU-) KAZUSA DNA RES INST.

XX (FUJI ) FUJISAWA PHARM CO LTD.

XX Miyoshi S, Zenkoh J, Satoh S, Nishimura S;

XX WPI; 2003-855124/80.

XX N-PSDB; ADD93812.

XX New polynucleotides encoding proteins useful for treating disorders  
XX associated with cell growth, division and death and hepatitis.

XX Claim 13; SEQ ID NO 2; 142pp; Japanese.

XX The invention relates to a splice variant of human HJ06972 (K1AA1079),  
CC designated AF06972 (ADD93815), and nucleic acids encoding it (ADD93814).  
CC Compared to the HJ06972 cDNA (ADD93812), AF06972 cDNA contains an  
CC additional 94 bp exon (ADD93816) between bases 4623-4624 of HJ06972,  
CC which results in a variant C-terminus (ADD93817) in AF06972. HJ06972 and  
CC its splice variant AF06972 are related to apoptosis-associated tyrosine  
CC kinase (AATK), a protein which is involved in apoptosis and the  
CC differentiation or proliferation of nerve cells, and are likely to have  
CC similar activity. The invention also relates to vectors and host cells  
CC comprising AF06972 DNA sequences, the recombinant production of AF06972  
CC polypeptides, antibodies specific for AF06972, an immunoassay method  
CC using the antibodies, a method for detecting apoptosis regulatory  
CC activity, a method for evaluating a compound for its ability to induce  
CC apoptosis or nerve cell differentiation or proliferation, and a  
CC transgenic non-human animal model in which apoptosis, or nerve cell  
CC division and growth has been modified. AF06972 polypeptides and  
CC polynucleotides, and related compounds and methods of the invention may  
CC be used in the treatment of disorders associated with apoptosis, nerve  
CC cell growth and division, and hepatitis. The present sequence represents  
CC the human HJ06972 protein.

XX SQ Sequence 1454 AA;

Query Match 6.5%; Score 152.5; DB 7; Length 1454;  
Best Local Similarity 21.4%; Pred. No. 0.0039;  
Matches 122; Conservative 75; Mismatches 199; Indels 183; Gaps 25;

Qy 3 LSKHQIQLSKPLSD---DSICGVYLKLSAFRLNEFNVAQTALRKLSONPSADERD 59  
Db 370 LPKPLEQ---PYSRWYELQFCWLSPEK---RPAEDVHRLTYLRLOSQDS--EVD 421

Qy 60 ALQEAQCLNKKWILSDLSVEQFSKTRDIELISWFAAQF-----LDDTTLSAANSLEW 113

Db 422 FEQQ-----WNALPNTNSRDSNNAAPFLDHPARDLGRMEEVLTVTETQGLSFEY 476

Qy 114 LADLSE-KHWD-----HLN-----PVLPVETLKSDDDKGEREQADAKVAFQL 157

Db 477 VWEAKKHDFERSGRHLDGLSYTSIFYPVEVESSLSDPGCKQ----- 522

Qy 158 VGDSEESILYAPVLQPLVGEVTFPD-----FQSAERKG---EISQLKSMITT 203  
Db 523 -DSDGQDVPLRVP-----GVVFDHNLVSGSDYIIQLEEKSGNLELDYPPALLTT 574  
Qy 204 TVAQERFAIQKMEKAKCVTQLDRLSALVSTKCHSLGSGSOSTNFGFAKSLLTRVENALVH 263  
Db 575 -----DMDNPRTGPELSQTLARLSE---LEESSTDEDFQSDTDPKDSLLPG 620  
Qy 264 LSGIKLAPKAKAVEQVEAESSVSEGELEPSHMDTKHIERIPMASEQAQ---TVSQHLH 319  
Db 621 DLHVTSGESPNNIFNDVDKSE---DLPSHQ--KIFDLMLNGVQADFKPAILSSSLD 674  
Qy 320 -----AGNLSELGNLNNMRDLAFHLREVSYDFRQSE 352  
Db 675 NPKSVITGHFEKEKPKRKFDESEPLCLSDNLHMQDNFDPLN-----VQELSENF--- 723  
Qy 353 PHSPISELEKAIKRWGYSLSPELLREMSSE-ONGDALSTIFNAAGLNHLD----- 401  
Db 724 -----LFLQEXNLLKGLSLSSKEHINDLQTELKNAAGTEAMLETSCRNSLDTELOFAENKP 778  
Qy 402 -----QVLL-----PEVSTPTVGIESPQTQPAKPSVSDPRSV 433  
Db 779 GMSLLQENVSTKGGDDTVMLTGDTLSLSTLQSSPEVQVPTSFETEETPRRVPDLSLPTQG 838  
Qy 434 EHVSTQSPVDTQSKQDQKQSSATSALS 462  
Db 839 E---TQPTCLDVIVPEDCLHQDISPDAVT 864  
RESULT 4  
AAW48845  
ID AAW48845 standard; protein; 1503 AA.  
XX  
AC AAW48845;  
XX 12-OCT-1998 (first entry)  
DE Human receptor tyrosine kinase LMR2\_h.  
XX Receptor tyrosine kinase; LMR2\_h; human; signal transduction; cancer;  
KW neurodegenerative disorder.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
FH Peptide 1..23  
FT /label= Sig\_peptide  
FT Domain 43..69  
FT /label= TMD  
FT /note= "transmembrane domain"  
FT Peptide 84..98  
FT /note= "immunogen"  
FT Peptide 163..166  
FT /note= "protein kinase conserved motif"  
FT Peptide 281..283  
FT /note= "protein kinase conserved motif"  
FT Peptide 329..332  
FT /note= "protein kinase conserved motif"  
FT Modified-site 409  
FT /label= O-phosphorylated  
FT Modified-site 477  
FT /label= O-phosphorylated  
FT Modified-site 552  
FT /label= O-phosphorylated  
FT Peptide 684..698  
FT /note= "immunogen"  
FT Modified-site 1032  
FT /label= O-phosphorylated  
FT Peptide 1098..1113  
FT /note= "immunogen"  
FT Modified-site 1100  
FT /label= O-phosphorylated  
FT Modified-site 1105











QY 388 LSTIFNAAGNLHDLQVLLPEVSTPTVGIESPOTPOAKPSVSDPRSVEHVSTQSPVDQTQS 447  
 DB 744 LRTV-----NMTWKNICREV--PTTLKECIOEPS---SVQTR-IAAHNVQKVLVSS 791  
 QY 448 KQDOKPOSSATSALS 462  
 DB 792 ASDIPVQSHRTSEIS 806

RESULT 10  
 AAU37669  
 ID AAU37669 standard; protein; 1179 AA.  
 AC AAU37669;  
 XX  
 XX 14-FEB-2002 (first entry)  
 XX  
 XX Streptococcus pneumoniae cellular proliferation protein #98.  
 DE  
 XX  
 XX Antisense; prokaryotic cellular proliferation protein; antibiotic;  
 KW antibacterial; drug design.  
 XX  
 XX Streptococcus pneumoniae.  
 OS  
 XX WO200170955-A2.  
 PN  
 XX 27-SEP-2001.  
 XX  
 XX 21-MAR-2001; 2001WO-US009180.  
 XX  
 XX 21-MAR-2000; 2000US-0131078P.  
 PR 23-MAY-2000; 2000US-0206848P.  
 PR 26-MAY-2000; 2000US-0207727P.  
 PR 23-OCT-2000; 2000US-0242578P.  
 PR 27-NOV-2000; 2000US-0253625P.  
 PR 22-DEC-2000; 2000US-0257931P.  
 PR 16-FEB-2001; 2001US-0269308P.  
 XX  
 XX (ELIT-) ELITRA PHARM INC.  
 XX  
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 PI  
 XX WPI: 2001-611495/70.  
 DR N-PSDB; AAS55528.  
 DR  
 XX New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids.  
 XX  
 XX Example 3; SEQ ID NO 13262; 511pp; English.  
 PS  
 XX The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the genes,  
 CC themselves and the encoded proteins. The prokaryotes used are Escherichia  
 CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,  
 CC pseudomonas aeruginosa and Enterococcus faecalis. The invention is also  
 CC useful for the identification of potential new targets for antibiotic  
 CC development. The antisense nucleic acids can also be used to identify  
 CC proteins used in proliferation, to express these proteins, and to obtain  
 CC antibodies capable of binding to the expressed proteins. The proteins can  
 CC be used to screen compounds in rational drug discovery programmes. The  
 CC antisense nucleic acid sequence is also useful to screen for homologous  
 CC nucleic acids which are required for cell proliferation in a wide variety  
 CC of organisms. The present sequence represents an essential prokaryotic  
 CC cellular proliferation protein. Note: The sequence data for this patent  
 CC did not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX Sequence 1179 AA;

Query Match 5.7%; Score 134.5; DB 4; Length 1179;

Best Local Similarity 21.6%; Pred. No. 0.092;  
 Matches 100; Conservative 80; Mismatches 178; Indels 105; Gaps 24;

QY 5 KHCIFQLKSPKSDSDSICGVYKLEKSAFPLNENVAQTALRKLSONFSADDERDALOE 64  
 DB 674 KPELSQLOKEIADE---ASLGSEEAALKTDQDMAALTELEAIK---SQEQARIQEQ 727  
 QY 65 CLNKWKILSDSLYEQFSKTTTRDIELISWFAVAAQFLDITLESAAANSLEWLADLSEKHW 124  
 DB 728 GLS-----LAYQOTSQVVELETL-WKLOEE-EIDRLSEG-----DWQAD-KEKQCES 772  
 QY 125 L-----NPVLPVETLKSDDDKGRERQADAKVKAFFQLVGSSESSILYAVLQPL 176  
 DB 773 LATIASERQNLEAEIEIKSNKNAIQERYQ-----NLOQEVAAQARLLKTEL 818  
 QY 177 VGEVTFDFQSAERKGEISQLKSMLETTTVAQERF--AIQFMENAKRCVCTQDRLSALVS 234  
 DB 819 CQOKR-YEVADIERLG-----KELDNLEIEEETQRMLOEKVDNLEKVDTE-----LLS 866  
 QY 235 TKCHSLGSGSTNF--GPAKSL--LTRVENALVHLSG-----IKLAPKAEAKTV 278  
 DB 867 QOAEESKTQKTNLQOGLIRKQFELDDIEGQDDIASHLDQARQOQNEEWIRKQTRAEAK-- 924  
 QY 279 EQEVAESSVSEGLPESHMDTKHIERIPWASEQAQTVSQHLHAGNLS--ELGNLNNMRDL 336  
 DB 925 KEKVSERL-----RHQONQLTQYQISYTEALEKAHELENLNLAEQEVODLEKAIRSL 977  
 QY 337 AFHLREVSDFRQSEPHSP:SFLLKKAIRWGYLSLPELLEMMSEQNGDA---LSTIFN 393  
 DB 978 GPNLEAIDQY---EEVHNRLDFL--NSQRDILSAKULLLETITENMDEVKRFKSTFE 1032  
 QY 394 AA-----GLNHLQVLLPEVSTPTVGIESPOTPOAK 424  
 DB 1033 AIRESFKVTFKQMGFGQAD-LITTEGDLTLAGVEISVQPPGK 1074

RESULT 11  
 AAU01107  
 ID AAU01107 standard; protein; 1179 AA.  
 XX  
 XX AAU01107;  
 AC  
 XX 02-OCT-2001 (first entry)  
 DT  
 XX CFE 110 protein sequence.  
 DE  
 XX Antibacterial; vaccine; gene therapy; bacterial cell wall viability; CFE;  
 KW CEG; Conserved Essential Gene; bacterial infection; antisense therapy;  
 XX antibiotic resistance.  
 XX  
 XX Streptococcus pneumoniae.  
 OS  
 XX WO200149721-A2.  
 XX  
 XX 12-JUL-2001.  
 PD  
 XX 29-DEC-2000; 2000WO-US035604.  
 XX  
 XX 30-DEC-1999; 99US-0174099P.  
 PR  
 XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 PA  
 XX Dougherty TJ, Pucci MJ, Dougherty BA, Davison DB, Brucoleri RE;  
 PI Thanassi JA;  
 PI  
 XX WPI: 2001-496721/54.  
 DR N-PSDB; AAH90806.  
 DR  
 XX Nucleic acids encoding conserved essential genes involved in bacterial  
 PT replication which are potential targets for the treatment of antibiotic  
 PT resistant bacterial infections.  
 XX  
 XX Claim 27; Page 365-369; 380pp; English.

XX The present invention relates to nucleic acids (AAH90701-AAH90918)  
CC encoding polypeptides (AAW01002-AAW0114), which are essential for the  
CC viability of a bacterial cell wall. The acronym CFE stands for "CEG For  
CC Expression", where CEG stands for "Conserved Essential Gene". The nucleic  
CC acids are useful for detecting the presence of proteins essential for the  
CC viability of a bacterial cell wall in samples such as cells, tissues,  
CC biological fluids, blood, serum, nose, ear or throat swabs with ligands,  
CC and for detecting corresponding target nucleic acid molecules with  
CC complementary sequences. The nucleic acids are also useful for  
CC determining whether a genomic nucleotide sequence of interest is  
CC essential for viability of a bacterial cell or whether it resides within  
CC an operon, by integrating an exogenous nucleotide sequence comprising a  
CC portion of an open reading frame of the genomic sequence of interest  
CC (comprising 200-500 base pairs) into the genomic sequence of interest  
CC which confers a selectable phenotype to the cell, and determining cell  
CC viability with a selection agent such as chloramphenicol. The nucleic  
CC acids and proteins are also useful as vaccines and for treating bacterial  
CC infections with gene therapy and antisense therapy. The nucleic acids  
CC also enable identification of targets suitable for the treatment of  
CC antibiotic resistant bacterial infections

XX  
SQ Sequence 1179 AA;  
Query Match 5.7%; Score 134.5; DB 4; Length 1179;  
Best Local Similarity 21.4%; Pred. No. 0.092;  
Matches 99; Conservative 80; Mismatches 180; Indels 103; Gaps 23;

QY 5 KHQIEQLSKPLSDSICGVYKLEKSAFRPLNEFNVAQTLRKLSQNSPADERDALQEA 64  
Db 674 KPELEQLQKEIAADE---ASLGSEEAALKTLQDQWALTERLEAIK---SQEQARIQEQ 727  
QY 65 CLNKWKILSDSLYEQFSKTRDIELISWFWAAQFLDITLESANSLWMLAD-----L 117  
Db 728 GLS-----LAYQTSQVVELETL-WKLQEE-EIDRLSEG-----DWQADKEKQESL 773  
QY 118 SEXHNDHNPVLVETILKSDDDKGRREQADAKVKAFFQLVGDSSESSILYAPVLQPLV 177  
Db 774 ATIASDKQNLAEIEEIKSNKNAIQERYQ-----NLOEEVAQRLKLTQ 819  
QY 178 GEVTFDFQSAEKGRHSIQLSKMLTTTVAQERF-AIQFMENAKRCVTQLDRLSALVST 235  
Db 820 GQKS-EVADIERLG-----KELDNLEIEQEEIQLMKEVDNLEKVDIE-----LLSQ 867  
QY 236 KCHSLGSQSNF--GPAKSL--LTRVENALVHLSG-----IKLAPAEAKTVE 279  
Db 868 QABESKTKTNLQGLIRKQFELDDIEGQLDDIASHLDQARQQNEWIRKQTRAEAK--K 925  
QY 280 QEVAESSVSGELPSPHMDTKHIERIPWASDAQTVSQHLHAGNLS--ELGNLNNMRDLA 337  
Db 926 EKVSERL-----RHLQNLQTDQYQISYTEALEKAHELENLNLAEQVQDLEKATRSLG 978  
QY 338 FHLREVSDFYRSEPHSPISFLEKAIKRWGYLSLPPELLREMMSEONGDA---LSTIFNA 394  
Db 979 PVNLEAIDQY---EEVHNRDLFL--NSQRDDILSAKNLLLETITENMDNVEKFRKSTFEA 1033  
QY 395 A-----CLNHLQDVLLPEVSTPTVGHIESPTQPAK 424  
Db 1034 IRESFKVTFKMFGGGQAD-LITTEGDLITAGVEISVQPPGK 1074

RESULT 12  
ABU46083  
ID ABU46083 standard; protein; 1179 AA.  
XX  
AC ABU46083;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by Prokaryotic essential gene #31610.  
XX  
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX

OS Streptococcus pneumoniae.  
XX  
FN W0200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
XX WPI; 2003-029926/02.  
DR N-PSDB; ACA4953.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
XX Claim 25; SEQ ID NO 74007; 1766pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs; or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 1179 AA;  
Query Match 5.7%; Score 134.5; DB 6; Length 1179;  
Best Local Similarity 21.6%; Pred. No. 0.092;  
Matches 100; Conservative 80; Mismatches 178; Indels 105; Gaps 24;

QY 5 KHQIEQLSKPLSDSICGVYKLEKSAFRPLNEFNVAQTLRKLSQNSPADERDALQEA 64  
Db 674 KPELEQLQKEIAADE---ASLGSEEAALKTLQDQWALTERLEAIK---SQEQARIQEQ 727  
QY 65 CLNKWKILSDSLYEQFSKTRDIELISWFWAAQFLDITLESANSLWMLADSEKHW 124  
Db 728 GLS-----LAYQTSQVVELETL-WKLQEE-EIDRLSEG-----DWQAD-KEKQES 772





CC with NOVX polypeptide. The DNA encoding the protein is useful in gene  
CC therapy for treating the disease or condition. In particular, the NOVX  
CC polypeptide or polynucleotide is useful for treating endocrine/  
CC metabolism-related diseases (e.g. obesity or diabetes), central nervous  
CC system disorders (e.g. Alzheimer's disease, Parkinson's disease,  
CC epilepsy, multiple sclerosis, schizophrenia or depression), autoimmune  
CC and inflammatory disorders (e.g. psoriasis, allergy, lupus erythematosus,  
CC asthma, inflammatory bowel disease, rheumatoid arthritis or  
CC osteoarthritis), cancers (e.g. colon, lung, liver, breast, ovarian,  
CC prostate or brain cancers, or melanoma), liver diseases (e.g. liver  
CC cirrhosis), lung diseases (emphysema or obstructive pulmonary disease),  
CC haemophilia, stroke, or infections (e.g. viral, bacterial or parasitic).  
CC These are also useful in developing powerful assay system for functional  
CC analysis of various human disorders, as well as in diagnostic  
CC applications, and for monitoring the effects of drugs during clinical  
CC trials. This is the amino acid sequence of a novel human NOV protein  
XX

SQ Sequence 1328 AA;

```
Query Match          5.5%; Score 129; DB 6; Length 1328;
Best Local Similarity 22.0%; Pred. No. 0.32;
Matches 108; Conservative 69; Mismatches 182; Indels 132; Gaps 23;

QY 3 LSKHQIEQLSKPLSDDSICGVYLKLEKSAFPLENEFVAQTALRKLSONPSADE---RD 59
DB 906 LKAHQVEVAQHNLKEASASQFELE-IVLKEKGNELKRLKLEMLKRESLSSKTQLLQD 964
QY 60 ALQEAACLNKWKI--LSDSLYEFQSKTTRDIELIS-----WFVAAQFLLOTTLE 105
DB 965 VQDENKLFKSIQELKQNYQACASSFPPEHELLKVISEREKEISGLWNE-----LDSLKD 1019
QY 106 SAANSLEWLADLSEKHWDHNLNPLVPVTLKSD--DDKGKREQ-----ADAKVKAF 155
DB 1020 AVEHQKKNNDLRKNWMEALASTEKMQLQDKVNTSKERQQQVEAVELEKEVLEKLF 1079
QY 156 QLVGDSERSSILYAPVLQPLVGEVTFPFDQSAERKGEISQLKSMLTITTTVAQERFAIQFK 215
DB 1080 PKV--SVFNSLSYGEWL-----HGFEKKAK-----ECMAGTSGSBEVKVLEHK 1120
QY 216 MENAKRCVTQDLRLSALVSTKCHSLGSGSTNFGPAKSLLTRVENALVHLSGIKLAPKAEA 275
DB 1121 LKEA-----DEMHTLLQLECEKY-----KSLVAETEGILQKQ-----1153
QY 276 KTVQGEVAESSVSGELPMDTHIERIPMASQOATVSOHLHAGNLSLGNLNNNRD 335
DB 1154 RSVQGEENKWKVYDE--SKTITQMGSSFTSSQ-----ELERLSENKD 1197
QY 336 LAFHLLREVSDYFRQSEPHSPISFLLEKA-----IRWGYLSLPELLREVMSEQN- 384
DB 1198 I-----ENLRREH--LEMELEKAEEMERTYVTEVRELKAQNLNETLKLRTQNE 1246
QY 385 -----GDALSTIFNAAGNLHDQVLLPEVSTPTVGIE-SPOTQAKPSVSDPRSVBEHVS 438
DB 1247 RQKVAGD-----LHKAQOSLELIQSKIKAAGDTTVIENSVDVSPETSESEKETMSVS--LN 1300
QY 439 QTSVPYDTQSKQ 449
DB 1301 QTV---TQLQQ 1308
```

Search completed: July 3, 2004, 02:44:29  
Job time : 81 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2004, 02:42:04 ; Search time 24 Seconds  
(without alignments)  
995.951 Million cell updates/sec

Title: US-10-780-347-2  
Perfect score: 2343  
Sequence: 1 MPLSKHQIEQLSKPLSDDSI.....DTQSKDQKPOSSATSLSW 463

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/iaa/5A-COMB.pep:\*  
2: /cgn2\_6/prodata/2/iaa/5B-COMB.pep:\*  
3: /cgn2\_6/prodata/2/iaa/5A-COMB.pep:\*  
4: /cgn2\_6/prodata/2/iaa/5B-COMB.pep:\*  
5: /cgn2\_6/prodata/2/iaa/5A-COMB.pep:\*  
6: /cgn2\_6/prodata/2/iaa/5B-COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	264.5	11.3	364	US-09-328-352-6641	Sequence 6641, Ap
2	197	8.4	681	US-09-328-352-6641	Sequence 25690, A
3	152.5	6.5	1503	US-08-976-255-14	Sequence 14, Appl
4	138	5.9	2008	US-09-091-501B-8	Sequence 8, Appl
5	137.5	5.9	357	US-09-252-991A-17293	Sequence 17293, A
6	126	5.4	610	US-09-336-447A-11	Sequence 11, Appl
7	123	5.2	949	US-09-252-991A-30776	Sequence 30776, A
8	120.5	5.1	999	US-08-770-301A-1	Sequence 1, Appl
9	120.5	5.1	999	US-09-175-581-1	Sequence 1, Appl
10	119.5	5.1	1066	US-09-541-782-8	Sequence 8, Appl
11	119.5	5.1	1066	US-09-723-820-8	Sequence 8, Appl
12	119.5	5.1	1066	US-10-270-085-8	Sequence 8, Appl
13	118.5	4.9	2468	US-09-976-594-321	Sequence 321, App
14	115.5	4.9	2468	US-09-976-594-726	Sequence 726, App
15	114.5	4.9	431	US-09-134-000C-5352	Sequence 5352, Ap
16	114.5	4.9	955	US-08-006-676B-1	Sequence 1, Appl
17	114.5	4.9	955	US-08-282-845-2	Sequence 2, Appl
18	114.5	4.9	955	US-08-428-41A-3	Sequence 3, Appl
19	114.5	4.9	955	PCR-US94-0034-1	Sequence 1, Appl
20	114	4.8	1939	US-09-310-187A-1	Sequence 1, Appl
21	112.5	4.8	1130	US-09-976-594-280	Sequence 280, App
22	110.5	4.7	381	US-09-216-393B-8	Sequence 8, Appl
23	110.5	4.7	675	US-08-317-522A-9	Sequence 9, Appl
24	110.5	4.7	675	US-08-439-818A-9	Sequence 9, Appl
25	110.5	4.7	675	US-08-751-965-9	Sequence 9, Appl
26	110.5	4.7	675	US-08-738-975-9	Sequence 9, Appl
27	110.5	4.7	675	US-08-728-626-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1  
US-09-328-352-6641  
; Sequence 6641, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 6641  
; LENGTH: 364  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-6641

Query Match 11.3%; Score 264.5; DB 4; Length 364;  
Best Local Similarity 21.5%; Pred. No. 1.7e-16;  
Matches 87; Conservative 76; Mismatches 152; Indels 89; Gaps 14;

Qy	8	IPQLSPKSDSDICGVYKLEKSAFPLRNENVAQTALRKLSONPSADERDALQEA	67
Db	5	ISELLKPIINDSLCG-----EDYSP---SNEFHEKKA--RTQDDLLDQGDWAERKQA	54
Qy	68	KWKILSDSLYEQFSKTRDIELISWFAAQFLDTTLESAANSLEWADLSEKWDHNP	127
Db	55	DMDFVAKSVSTLLIEKTDIRLLTWIBAWTHLN-GPEGVVKGITLTHMLNQWQDIHP	113
Qy	128	VLPVETLKSDDDKGRERQADAKVAFQLYGDSSESLVAPVL--OLPLVGEVTFP--	183
Db	114	II-----EDDD-----LDQRIQLQLNQ-----LPMLKKVPLTNTAPYNL	152
Qy	184	-----DFQSAERKGIISQLKMLTTVAQERFAIQKMNKRCVQTQ	225
Db	153	LDYDNFLYHENIRRKQTBESVQSQSPSELEQDQAFNT-----SKTFQSYNTQFNSV	206
Qy	226	LDRLSALVETKCHSLGSGOSTNFGPAKSLITLVEVAL-----VHLSGKILAPKAKTVEQ	280
Db	207	LTEWNVKLTDLHGLDGLDSPFAAIDSAFETIHLTKIYKAEAFGTGLAPSOQAAY--	264
Qy	281	EVAESVSGELPUSHMDTKNHERIPMASEQATQVSQHUHAGNLSELNLNNDRLAFHL	340
Db	265	-----ITTFSENVQPVQIVSQDQPMFQFQACT---HL-----ANREQAMKV	302
Qy	341	LRVSDYFRQSPHSPISFLLEKATRWGYSILPELLREMMSEON	384
Db	303	LQEIADYFOANEPHSPVSMLOKTIKNSQMPHEWLAQVINKDEH	346





## RESULT 4

US-09-091-501B-8  
 ; Sequence 8, Application US/09091501B  
 ; Patent No. 6518413  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tinsley, Jonathon M  
 ; APPLICANT: Davies, Kay E  
 ; TITLE OF INVENTION: Trophin gene expression  
 ; FILE REFERENCE: 620-42  
 ; CURRENT APPLICATION NUMBER: US/09/091,501B  
 ; CURRENT FILING DATE: 1998-06-18  
 ; PRIOR APPLICATION NUMBER: PCT/GB96/03156  
 ; PRIOR FILING DATE: 1996-12-19  
 ; PRIOR APPLICATION NUMBER: GB 9525962.8  
 ; PRIOR FILING DATE: 1995-12-19  
 ; PRIOR APPLICATION NUMBER: GB 9615797.9  
 ; PRIOR FILING DATE: 1996-07-26  
 ; PRIOR APPLICATION NUMBER: GB 9622174.2  
 ; PRIOR FILING DATE: 1996-10-24  
 ; NUMBER OF SEQ ID NOS: 15  
 ; SOFTWARE: Patent in Ver. 2.1  
 ; SEQ ID NO 8  
 ; LENGTH: 2008  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURES:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (239)... (250)  
 ; OTHER INFORMATION: Description of Artificial Sequence: Chimeric; Xaa = Unknown

Query Match 5.8%; Score 138; DB 4; Length 2008;  
 Best Local Similarity 21.2%; Pred. No. 0.0023;  
 Matches 105; Conservative 72; Mismatches 172; Indels 146; Gaps 19;  
 Qy 1 MPKSHQIGDLSKPLSDSDICGVYKLEKSAFPLRNFVAQTALRKLSQNPSPADDERA 60  
 Db 420 MELQKQLQSLWLA-----LTERIQKQSPPLGDDLPS 455  
 Qy 61 LOEACLNKWLSDSLVQFSKTRDIEL-----SMFVAAQFLDITLESANSLWLA 115  
 Db 456 LQK-LLOEHSKLNQDLAEQVKNSITHWVIVDENGSGATALLDQKLGERTWAVC 514  
 Qy 116 DLSEKHWDHNLNPLVETLKSDDDKGKERQADAKVAFQVLGVDSSESSILYAPVLQIP 175  
 Db 515 RWTERRWNRL-----QETSILWQELLBEQ 538  
 Qy 176 LVGEVTFDFQSAERKEISOLKSMULTTTVAQERFAIQKMNKAKCVTQDLRLSAL--- 232  
 Db 539 CLEAWLTEREALNKVQTNFQKQKLSVSVRRLLAKEDMEMKR--OTLDQSEIQD 596  
 Qy 233 -----VSTKCHSLGSGSTNFGPAKSLITRVENALVHLSGILAPKAAEAKTYEQEVA 283  
 Db 597 VGQLLSNPKASKXKNSDSELTQRW--DSLVSQLEDS-----SNQVTQAVA 640  
 Qy 284 ESSVSEGEPLSHMDTKHIERIPNASEAQ-----TVSQH-LHAGNLSELGNLNNRDL 336  
 Db 641 KLGNSQIPQDLLETTHVREKGMVKPKQELPPLTKAHHMKQSTTELG----- 691  
 Qy 337 AFHLRLREVSDFRQSEPH-----SPISFLEKAIKRWGLSLPELLREMMSEQNGDA 387  
 Db 692 --ENLQELRLDTQEMEVAEKLKWLNRTEMLSDKS-----LSLPE--RDKISE---S 738  
 Qy 388 LSTIFNAGLNHLDOVLLPVSPTVGIESPOTPOAKPSVSDPRSEVHSQTSVDVDTQS 447  
 Db 739 LRTV-----NMTWKNICREV--FTTLKECICQPS---SVSQTR-IAAHNVQKXVLVSS 786  
 Qy 448 KQDQKPOSSATSALS 462  
 Db 787 ASDIPVQSHRTSBS 801

## RESULT 5

US-09-252-991A-17293  
 ; Sequence 17293, Application US/09252991A  
 ; Patent No. 8551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 17293  
 ; LENGTH: 357  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; US-09-252-991A-17293

Query Match 5.9%; Score 137.5; DB 4; Length 357;  
 Best Local Similarity 21.6%; Pred. No. 0.00016;  
 Matches 91; Conservative 67; Mismatches 145; Indels 119; Gaps 18;  
 Qy 15 LSDSDICGVYKLEKSAFPLRNFVAQTALRKLSQ-NPSADERDALQEAELNKKWILS 73  
 Db 24 VSPDPCGDDL-----EYDAFLELERTAQOPEROMGDVLPFAEPPEWPRVR 71  
 Qy 74 DSVLEQFSKTRDIELISWFAAQFLDITLESANSLWLAADLSEKHWHDHNLNPLVET 133  
 Db 72 ALASELFGK-SKDLRVANLLQSNVALD-GLDGLADGLLVRELLQYVGVPL----- 124  
 Qy 134 LKSDDDKGREREQADAKVAFQVLGVDSSESSILYAPVLQ-----PLICEVTP--FDQS 187  
 Db 125 LDADDD-----NPTFRINALTGLVAE-----PLQLVWALPLVRSRAGPQVNDRA 170  
 Qy 188 AERKGISQLKSMULTTTVAQERFAIQKMNKAKCVTQDLRLSALVSTKCHSLGSGSTNF 247  
 Db 171 ALNAAQLQRFAS--ETLSPEQIAGAFADADA-----DALAATRALEGAQEAH 217  
 Qy 248 GFAKSLITRVENA-----LVHL-----SGIKLAPKAAKTVQEVAES 285  
 Db 218 AIBSGVAERYGSAQGLDLQLRLQALQVFDLYGPOGAGEFLAPGAAEAADEOGGA-A 276  
 Qy 286 SVSEGEPLSHMDTKHIERIPNASEAQTVSQHLHAGNLSELGNLNNRDLAFHLRLREVS 345  
 Db 277 PVAAVAAPA-----PRAS-----GEIA-----NREDEVLRQLDRLL 306  
 Qy 346 DYPRQSEPHSPISFLEKAIKRWGLSLPELLREMMSEQNGDALSTIFNAGLNHLDOVLL 405  
 Db 307 EYVREHPSFPVLLKRAKTLVTADFABIVRNLPD-----GTSQFETLGRG 353  
 Qy 406 PE 407  
 Db 354 PE 355

## RESULT 6

US-09-336-447A-11  
 ; Sequence 11, Application US/09336447A  
 ; Patent No. 6310190  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HANSEN, ERIC J.  
 ; APPLICANT: ABEI, CHRISTOPH  
 ; APPLICANT: COPE, LESLIE D.  
 ; APPLICANT: MACIVER, ISOBEL  
 ; APPLICANT: FISKE, MICHAEL J.  
 ; APPLICANT: FREDENBURG, ROSS A.  
 ; TITLE OF INVENTION: USP41 AND USP42 ANTIGENS OF MORAXELLA CATARRHALIS  
 ; FILE REFERENCE: AMCY:024

;; CURRENT APPLICATION NUMBER: US/09/336.447A  
;; CURRENT FILING DATE: 1999-06-21  
;; NUMBER OF SEQ ID NOS: 98  
;; SOFTWARE: Patent in Ver. 2.1  
;; SEQ ID NO 11  
;; LENGTH: 610  
;; TYPE: PRT  
;; ORGANISM: Moraxella catarrhalis  
US-09-336-447A-11

Query Match 5.4%; Score 126; DB 4; Length 610;  
Best Local Similarity 20.8%; Pred. No. 0.0045;  
Matches 82; Conservative 71; Mismatches 166; Indels 76; Gaps 15;

QY 43 QTALRKLSONPSADERDALO--EACLNKWKIL-----SDSLYQFQSKTTDIEL--- 89  
DB 37 QDSISKLVQ-----DDITLQDQOQXMKYLLNLQNLANTLITDELNNVKNNTSIEALGD 92  
QY 90 -ISWFAAQFLDITLESAAANSLEWLADLSEKHWDHNPVLPVETILKSDDDKEREQAD 148  
DB 93 EIGWLENDIADLEBGEVETLKNQNTLIEKDEH-----DRLIAQNOAD 135  
QY 149 AK-----VYAFQLVGD--SERSSILYAPVLQPLVGEVTFPDPQSAERKEISOLKSM 200  
DB 136 IOTLENNVVEELNLSGLRLDQEADI-----AKNASIEELYDPDNEVAERIGEIHAYTEE 191  
QY 201 LTTTVAQERFAIQKVENAKRCVQTDRLSALVSTKCHSL-----GSQSTNFGFAKSLLTR 256  
DB 192 VNKT-----LENLITNSVKNNTNDIKNKAADINNIHIELAQOQDQSHSSDIKTLKN 244  
QY 257 VENALVHLSGKILAPKAEAKTIVEQVAESVSGELPSPMDTKHIERIPMAEQATVSVQ 316  
DB 245 VEGLLESLGHLIDQKADL--TKDIAKESNVBEGLL--DLSGRLLDQKADLTAKDIAKES 301  
QY 317 HLHAGNLSLGNLNNNRDLA-----FHLREVSDFROSEPHSPISFLLKKAIRWG 368  
DB 302 NVEGLLDLSGLRLDQKADIAQNTQIDILAYNELQOYAKQTEAIDA--LNKASSEN 359  
QY 369 YLSLPEL-----LREMMSQNGDALSTIFNAAGLN 398  
DB 360 TONIEDLAAYNELQDAYAKQTEAIDALNKASSEN 394

RESULT 7  
US-09-252-991A-30776  
; Sequence 30776, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 30776  
; LENGTH: 949  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-30776

Query Match 5.2%; Score 123; DB 4; Length 949;  
Best Local Similarity 23.8%; Pred. No. 0.018;  
Matches 108; Conservative 69; Mismatches 203; Indels 74; Gaps 20;

QY 5 KHOIQLKSLDDSGICGVYGLKLSAFRPRNEFNVAQTALR--KLSONPSADERDALQ 62  
DB 490 KERLEKELKQLKTQSQVAADRAASQAQAEKLYQDVLDQAQKALEDFRRTQTTLTAPEPKIA 549

QY 63 -----EACLNKWKILSDSLYQFQSKTTDIELISWFAAQFLDITLESAAANSLEWL-AD 116  
DB 550 ELAQLEASQDELKRSSDAFAERVQOLSARLQVAFQGLDLEAKERTULEDALRRQLLPAD 609  
QY 117 L-----SEKHWDHNPVLPVETILKSDDDKEREQADAKVKAFQOLVGDSEESSILYAP 170  
DB 610 LPFGTFPNEPVDSDSLNLLPL--LNDYQDTQALQCRIDGQIEA-----LYA- 653  
QY 171 VLQPLVGEVTFPDPQSAERKEISOLKSMLETTTVAQERFAIQKVENAKR-CVTQDLRL 229  
DB 654 --QVRLKGVAKFDSDDTERR-----LKLINAYAHQDEAL--TLAKARRAAVTDIAR- 703  
QY 230 SALVSTKCHSLSGSOSTNFGFAKSLLTRVEN-----ALVHLSGILKAPKAEA-KTVEQEV-A 283  
DB 704 -----TURNIRSDYDNLHQLALFNREINKRQVSNLQSFRIVLAPNCKDKKHIDQIH 757  
QY 284 ESSVSEGEPLSPMDTKHIERIPMAEQATVSVQHLHAGNLSLGNLNNNRDLAFHLLRE 343  
DB 758 AGQVEEGENLSVFDLTQSAEQDAKNEEAKVEYLAELVAANHNLG-----LKDL-FELAFE 811  
QY 344 VSDYFRQSEPHSPI-----SFLEKAIKRWGILSLPELLREMMSQNGDALSTIF--NA 394  
DB 812 ITKVGQPIVHTDIDGAASNGTMTIKALTNMYL-----LLHLMRDREQAGRVLPYPYLDEA 867  
QY 395 AGLNHLQVLLPEVSTPTVGIESPQTPQAKPSVS 428  
DB 868 ADIDERNQALLETSLQLGFV--PILASVKEPVS 899

RESULT 8  
US-08-770-301A-1  
; Sequence 1, Application US/08770301A  
; Patent No. 5948637  
; GENERAL INFORMATION:  
; APPLICANT: IKEDA, JUN  
; APPLICANT: KANEDA, SUMIKO  
; APPLICANT: YANAGI, HIDEKI  
; APPLICANT: MATSUMOTO, MASAYASU  
; APPLICANT: YURA, TAKASHI  
; TITLE OF INVENTION: NOVEL STRESS PROTEINS  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH  
; STREET: PO BOX 747  
; CITY: FALLS CHURCH  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/770,301A  
; FILING DATE: 20-DEC-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURPHY JR, GERALD M  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 1422-287  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)-205-8000  
; TELEFAX: (703)-205-8050  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 999 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-770-301A-1

Query Match 5.1%; Score 120.5; DB 2; Length 999;

```
Best Local Similarity 22.2%; Pred. No. 0.033;
Matches 89; Conservative 63; Mismatches 122; Indels 127; Gaps 19;

Qy 81 SKTTTRDIELISWFAAQFLDITLESAAANSLEWADLSEKHWDLNPLVPLVETLKSDDDK 140
Db 690 ARKRMRVEEIGVELVLDLPDLPEDKLAQSVQKLQDL-----TLR--DLE 732
Qy 141 GKREQADAKVKAFFQVLGDSSESSILYAPVLQPLVGEVTFDFQSAERKEISQLKSM 200
Db 733 KQREKAANSLEAFI-----FETQDKLYQPEYQ-----EVS--TEEQREIEISGKLSAASTW 781
Qy 201 L-----TTTVAQERFA-----IQKMNNAKRCVTQDLRLSALVSTKCHSLGSOST 245
Db 782 LEDEGVGATTVMLEKLAELRKLCOGLFRVEERKKWP-----ERLSALDNLNHS----- 832
Qy 246 NFGFAKSLLRVENALVHLSGILAPKAAE--KTVEQEAESSVSEGEPLPSHMDTKHIER 303
Db 833 -----SMFLKGARLIPMDQIFTEVEMTTLEKVINETWAKNATLAEQAK 877
Qy 304 IPWASEQAQTVSOHLHAGNLSELGNLNNNRDLAFHLLREVSDFYFQSEPHSPISFLLEK 363
Db 878 LP-ATEKPVLLSKDIEAKMA-----LDREV-----QYLLNK 908
Qy 364 AIRWGYLSLPELLREMSQNGDALSTIFNAAGLNHLDOVLLPEVSTPTVGIESPOTPOA 423
Db 909 A-----KFTKP---RPRPKDKNGTRAEPPLNASASDQGEKVIPP-----AGQTEDA 951
Qy 424 KPSVSDPRSVEEHVSQTSPTDQ-----SKDQKPOSS 456
Db 952 EP-ISEPEKVE---TGSEPGDTEPLELGGPGAPEQKEQST 988

RESULT 9
US-09-175-581-1
; Sequence 1, Application US/09175581
; Patent No. 6034232
; GENERAL INFORMATION:
; APPLICANT: IKEDA, JUN
; APPLICANT: KANEDA, SUMIKO
; APPLICANT: YANAGI, HIDEKI
; APPLICANT: MATSUMOTO, MASAYASU
; APPLICANT: YURA, TAKASHI
; TITLE OF INVENTION: NOVEL STRESS PROTEINS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/175,581
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/770,301
; FILING DATE: 20-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR, GERALD M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1422-287
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)-205-8000
; TELEFAX: (703)-205-8050
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 999 amino acids
```

```
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-175-581-1

Query Match 5.1%; Score 120.5; DB 3; Length 999;
Best Local Similarity 22.2%; Pred. No. 0.033;
Matches 89; Conservative 63; Mismatches 122; Indels 127; Gaps 19;

Qy 81 SKTTTRDIELISWFAAQFLDITLESAAANSLEWADLSEKHWDLNPLVPLVETLKSDDDK 140
Db 690 ARKRMRVEEIGVELVLDLPDLPEDKLAQSVQKLQDL-----TLR--DLE 732
Qy 141 GKREQADAKVKAFFQVLGDSSESSILYAPVLQPLVGEVTFDFQSAERKEISQLKSM 200
Db 733 KQREKAANSLEAFI-----FETQDKLYQPEYQ-----EVS--TEEQREIEISGKLSAASTW 781
Qy 201 L-----TTTVAQERFA-----IQKMNNAKRCVTQDLRLSALVSTKCHSLGSOST 245
Db 782 LEDEGVGATTVMLEKLAELRKLCOGLFRVEERKKWP-----ERLSALDNLNHS----- 832
Qy 246 NFGFAKSLLRVENALVHLSGILAPKAAE--KTVEQEAESSVSEGEPLPSHMDTKHIER 303
Db 833 -----SMFLKGARLIPMDQIFTEVEMTTLEKVINETWAKNATLAEQAK 877
Qy 304 IPWASEQAQTVSOHLHAGNLSELGNLNNNRDLAFHLLREVSDFYFQSEPHSPISFLLEK 363
Db 878 LP-ATEKPVLLSKDIEAKMA-----LDREV-----QYLLNK 908
Qy 364 AIRWGYLSLPELLREMSQNGDALSTIFNAAGLNHLDOVLLPEVSTPTVGIESPOTPOA 423
Db 909 A-----KFTKP---RPRPKDKNGTRAEPPLNASASDQGEKVIPP-----AGQTEDA 951
Qy 424 KPSVSDPRSVEEHVSQTSPTDQ-----SKDQKPOSS 456
Db 952 EP-ISEPEKVE---TGSEPGDTEPLELGGPGAPEQKEQST 988

RESULT 10
US-09-541-782-8
; Sequence 8, Application US/09541782
; Patent No. 6284480
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Seraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/541,782
; CURRENT FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1066
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-541-782-8

Query Match 5.1%; Score 119.5; DB 3; Length 1066;
Best Local Similarity 19.1%; Pred. No. 0.046;
Matches 96; Conservative 78; Mismatches 161; Indels 167; Gaps 21;

Qy 5 KHQIEQLSKPLSDSDSIC-----GVYLKLESAFRPLRNFENVAQTALRKLSONPSADER- 58
Db 373 KEYTEBIDK-LKEDLMAARDKNGIYLA-----EETGYEITLKLESQNRLENEKM 420
Qy 59 ---DALQEACLNKWKILSD---SLYEQ-----FSKTTTRDIE 88
Db 421 LLLKALKDELQNKKEIFSEVMSLVKTQELKYTEENLNTKGTLLTKKVLTKTKRXYK 480
Qy 89 LISWFAAQFLDITLESAAANSLEWADLSEKHWDLNPLV-----VETLKSDDDKGE 143
Db 481 EKKELVASHMKTEQVLTQAEILAAADLATDTHLGHGTIERREIRDELKIRRSQDFKD 540
```

QY 144 REQADAKVKAFFQVLDSESSILYAPVLQPLVGEVTFDFQSAERKEISQLKSLMTT 203  
 Db 541 RMQDN-----LEMIG-----GSLNLYQDOQAALKEQLSQ--EMVNS 574  
 QY 204 TVAQERFAIOFKMNAKRCVTOLDRLSALVSTKCHSLGSGSTNFGFAKSLTRVENALVH 263  
 Db 575 SYVSQRLAL-----NSSKSIEMLKEMCA-----QSLQDQ-----TNLHNKLI- 611  
 QY 264 LSGIKLAPKABAKTVEQVAESSVSEGEPLPSHMDTKHIERIPMASEQAQTSQHLHAGNL 323  
 Db 612 -----GEVMKISDQHSQAFVA--KLMEQMQOQQL-----LMSKEIQTNLQVIEENNQ 656  
 QY 324 SELGNLNNNRDLAPHL---LREVSDFRQSEPHSPISFLLEKAIKRWGYSLP----- 373  
 Db 657 RHKAMLDQMOKFATIIDSSLSQSVVEHAQ-----MHKLEQLGAMSLPDAEELQN 707  
 QY 374 ---ELLREMMSEONGDAL-----STIFNAGLNHLDOVLLPE----- 407  
 Db 708 LOEELANERALAQQEDALLESMQMOMEQIKNLRKSNKISMSVHLNKWESRLTRNHRIDD 767  
 QY 408 -----VSTPTVGIESPOTPOAK 424  
 Db 768 IKSGIQDYQKLGIEASQSAQAE 789

## RESULT 11

US-09-723-820-8  
 ; Sequence 8, Application US/09723820  
 ; Patent No. 6468760  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Nislow, Corey  
 ; APPLICANT: Sakowicz, Roman  
 ; APPLICANT: Beraud, Christophe  
 ; TITLE OF INVENTION: Antifungal Assay  
 ; FILE REFERENCE: 1015  
 ; CURRENT APPLICATION NUMBER: US/09/723,820  
 ; PRIOR FILING DATE: 2000-11-28  
 ; PRIOR APPLICATION NUMBER: 09/541,782  
 ; PRIOR FILING DATE: 2000-04-03  
 ; NUMBER OF SEQ ID NOS: 10  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 8  
 ; LENGTH: 1066  
 ; TYPE: PRT  
 ; ORGANISM: Drosophila melanogaster  
 US-09-723-820-8

Query Match 5.1%; Score 119.5; DB 4; Length 1066;  
 Best Local Similarity 19.1%; Pred. No. 0.046;  
 Matches 96; Conservative 78; Mismatches 161; Indels 167; Gaps 21;  
 QY 5 KHOIEQLSKPLSDSIC-----GVYKLEKSAFPLRNEFNVAQTALRKLSONPSADER- 58  
 Db 373 KEYTEEDK-LKRDLMARDKNGIYLA-----ETYGEITLKESQRELNEKM 420  
 QY 59 ---DALQEAACLNKWKILSD---SLYEQ-----FSKTRDIE 88  
 Db 421 LLLKALKDELQNKKEIFSEVMSLVETQELKKEENLNTKGTLLTKKVLTKRYK 480  
 QY 89 LISNFWAAQFLDITLESAAANSLEWLADLSEKHWDHNPVLP-----VETLKSDDDKGE 143  
 Db 481 EKKELVASHMKTQEVLTQAEILAAADLATDTHQLHGHTIERRELDEKIRSCDFKD 540  
 QY 144 REQADAKVKAFFQVLDSESSILYAPVLQPLVGEVTFDFQSAERKEISQLKSLMTT 203  
 Db 541 RMQDN-----LEMIG-----GSLNLYQDOQAALKEQLSQ--EMVNS 574  
 QY 204 TVAQERFAIOFKMNAKRCVTOLDRLSALVSTKCHSLGSGSTNFGFAKSLTRVENALVH 263  
 Db 575 SYVSQRLAL-----NSSKSIEMLKEMCA-----QSLQDQ-----TNLHNKLI- 611  
 QY 264 LSGIKLAPKABAKTVEQVAESSVSEGEPLPSHMDTKHIERIPMASEQAQTSQHLHAGNL 323  
 Db 612 -----GEVMKISDQHSQAFVA--KLMEQMQOQQL-----LMSKEIQTNLQVIEENNQ 656  
 QY 324 SELGNLNNNRDLAPHL---LREVSDFRQSEPHSPISFLLEKAIKRWGYSLP----- 373  
 Db 575 SYVSQRLAL-----NSSKSIEMLKEMCA-----QSLQDQ-----TNLHNKLI- 611  
 QY 264 LSGIKLAPKABAKTVEQVAESSVSEGEPLPSHMDTKHIERIPMASEQAQTSQHLHAGNL 323

Db 612 -----CEVMKISDQHSQAFVA--KLMEQMQOQQL-----LMSKEIQTNLQVIEENNQ 656  
 QY 324 SELGNLNNNRDLAPHL---LREVSDFRQSEPHSPISFLLEKAIKRWGYSLP----- 373  
 Db 657 RHKAMLDQMOKFATIIDSSLSQSVVEHAQ-----MHKLEQLGAMSLPDAEELQN 707  
 QY 374 ---ELLREMMSEONGDAL-----STIFNAGLNHLDOVLLPE----- 407  
 Db 708 LOEELANERALAQQEDALLESMQMOMEQIKNLRKSNKISMSVHLNKWESRLTRNHRIDD 767  
 QY 408 -----VSTPTVGIESPOTPOAK 424  
 Db 768 IKSGIQDYQKLGIEASQSAQAE 789

## RESULT 12

US-10-270-085-8  
 ; Sequence 8, Application US/10270085  
 ; Patent No. 6627408  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Nislow, Corey  
 ; APPLICANT: Sakowicz, Roman  
 ; APPLICANT: Beraud, Christophe  
 ; TITLE OF INVENTION: Antifungal Assay  
 ; FILE REFERENCE: 1015  
 ; CURRENT APPLICATION NUMBER: US/10/270,085  
 ; CURRENT FILING DATE: 2002-10-11  
 ; PRIOR APPLICATION NUMBER: US/09/723,820  
 ; PRIOR FILING DATE: 2000-11-28  
 ; PRIOR APPLICATION NUMBER: 09/541,782  
 ; PRIOR FILING DATE: 2000-04-03  
 ; NUMBER OF SEQ ID NOS: 10  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 8  
 ; LENGTH: 1066  
 ; TYPE: PRT  
 ; ORGANISM: Drosophila melanogaster  
 US-10-270-085-8

Query Match 5.1%; Score 119.5; DB 4; Length 1066;  
 Best Local Similarity 19.1%; Pred. No. 0.046;  
 Matches 96; Conservative 78; Mismatches 161; Indels 167; Gaps 21;  
 QY 5 KHOIEQLSKPLSDSIC-----GVYKLEKSAFPLRNEFNVAQTALRKLSONPSADER- 58  
 Db 373 KEYTEEDK-LKRDLMARDKNGIYLA-----ETYGEITLKESQRELNEKM 420  
 QY 59 ---DALQEAACLNKWKILSD---SLYEQ-----FSKTRDIE 88  
 Db 421 LLLKALKDELQNKKEIFSEVMSLVETQELKKEENLNTKGTLLTKKVLTKRYK 480  
 QY 89 LISNFWAAQFLDITLESAAANSLEWLADLSEKHWDHNPVLP-----VETLKSDDDKGE 143  
 Db 481 EKKELVASHMKTQEVLTQAEILAAADLATDTHQLHGHTIERRELDEKIRSCDFKD 540  
 QY 144 REQADAKVKAFFQVLDSESSILYAPVLQPLVGEVTFDFQSAERKEISQLKSLMTT 203  
 Db 541 RMQDN-----LEMIG-----GSLNLYQDOQAALKEQLSQ--EMVNS 574  
 QY 204 TVAQERFAIOFKMNAKRCVTOLDRLSALVSTKCHSLGSGSTNFGFAKSLTRVENALVH 263  
 Db 575 SYVSQRLAL-----NSSKSIEMLKEMCA-----QSLQDQ-----TNLHNKLI- 611  
 QY 264 LSGIKLAPKABAKTVEQVAESSVSEGEPLPSHMDTKHIERIPMASEQAQTSQHLHAGNL 323  
 Db 612 -----GEVMKISDQHSQAFVA--KLMEQMQOQQL-----LMSKEIQTNLQVIEENNQ 656  
 QY 324 SELGNLNNNRDLAPHL---LREVSDFRQSEPHSPISFLLEKAIKRWGYSLP----- 373  
 Db 657 RHKAMLDQMOKFATIIDSSLSQSVVEHAQ-----MHKLEQLGAMSLPDAEELQN 707  
 QY 374 ---ELLREMMSEONGDAL-----STIFNAGLNHLDOVLLPE----- 407

Db 708 LQELANERALAOEDALLSEMMQVEQIKNLRKNSISVHLNKMESRLTRNHRIDD 767  
QY 408 -----VSTETVIGIESPQTQAK 424  
Db 768 IKSGIQDYOKLGIASQSAQAE 789

RESULT 13  
US-09-976-594-321  
; Sequence 321, Application US/09976594  
; Patent No. 6673549  
; GENERAL INFORMATION:  
; APPLICANT: Furness, Michael  
; APPLICANT: Buchbinder, Jenny  
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
; FILE REFERENCE: PA-0041 US  
; CURRENT APPLICATION NUMBER: US/09/976,594  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: 60/240,409  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 1143  
; SOFTWARE: PERL Program  
; SEQ ID NO 321  
; LENGTH: 806  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. 6673549 3768043CD1  
US-09-976-594-321

Query Match 5.1%; Score 118.5; DB 4; Length 806;  
Best Local Similarity 20.7%; Pred. No. 0.036;  
Matches 110; Conservative 80; Mismatches 217; Indels 125; Gaps 20;

QY 4 SKH-----QIEQLSKPLSDDS-----ICGYLKLEKS---AFPLRNEFN 40  
Db 158 SKHSQDELMAIDSKISQSLDESKSLKSQIAEAKIICKTFKMSERRAIRAIKDALNENS 217

QY 41 VAOTALKILQSPADDERDALQEAELNKNWILSDLSYEQFSKTRDIELISWFAQAFL 100  
Db 218 QLQTHKQLFOQAEVWKVEVSE--LNKQKI----- 246

QY 101 DTTLESANSLWLAIDLSERHDLNPVLP-----VETLSDDDKKGKROQAOKVK 152  
Db 247 --TFEDSKVHAEQVLNDKENHIKTLTGHLPMKQQAALVEEDTDDNLEVNSESENG 304

QY 153 AFFOLVGDSESSILYAPVLQ---PLVGEVTFDFQSAE---RKGEISO-LKSMLTTV 205  
Db 305 AYLDNPPKGAJ.KKLIHAAKLNASLKTLEGERNQIYIQLSEVDKTEBELTIEHIKNLQTOA 364

QY 206 AQERFAIQFMENAK-----RCVTLQDLRLSALVSTKCHSLGSGSTNFGFAK-SILLTRVEN 259  
Db 365 SLQSETHFENENQKLOQKLKVMTELYQENEM---KLHRKLTVEENVYRLEKEEKLKSKVDE 421

QY 260 ALVHLSGTLKAPKAETVQEVAEBSVS-BGELPSHMDTKHIERIPMAEQACTVSOHL 318  
Db 422 KISHATELETYRKAQLEBEELERTHSYQGGIISHEKKAHDNW--LAARNAR----- 474

QY 319 HAGNLSELGNLNNMR-----DLAFHLLRE-----VSDYPRQSEPHSPISFILLEKA 364  
Db 475 ---NUNLURKENAHNRQKLTETELKFELEKDPALDVPNTAFGREHSPYCPSPGLWPSS 531

QY 365 IRWGYLSLPELLREM-----SQNGDALS-TIFNAGLNHLDDQVLLPEVST 410  
Db 532 ETRAFLSPTLLLEGPLRLSPLLPGGGGRSGRPGNPLDHOITNBERGESSCDRLDPRAP 591

QY 411 PTVGLIESQTOAKPSVSDPRSVEHVSTQSPVDTQSKQDQKPOSSATSALS 462  
Db 592 SDTGLSLP-----PWDQDRMMFFPPGQSYDPSALPQRODRFCNSGRLS 637

RESULT 14  
US-09-976-594-726  
; Sequence 726, Application US/09976594  
; Patent No. 6673549  
; GENERAL INFORMATION:  
; APPLICANT: Furness, Michael  
; APPLICANT: Buchbinder, Jenny  
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
; FILE REFERENCE: PA-0041 US  
; CURRENT APPLICATION NUMBER: US/09/976,594  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: 60/240,409  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 1143  
; SOFTWARE: PERL Program  
; SEQ ID NO 726  
; LENGTH: 2468  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. 6673549 4295277CD1  
US-09-976-594-726

Query Match 4.9%; Score 115.5; DB 4; Length 2468;  
Best Local Similarity 21.4%; Pred. No. 0.42;  
Matches 119; Conservative 78; Mismatches 199; Indels 161; Gaps 27;

QY 2 PLSKHQIEQLSKPLSDDSICGYLKLEKSAPRPLRNEFNVAQTALRKLQSN----- 52  
Db 1209 PPSMNEEDKFSRSLRDVACS-----EVKASTLLDKDSIGAVSSEKVPSPKSPSLSPSP 1263

QY 53 PSADDERDALQEAELNKNWILSDLSYEQFSKTRDIEL-----ISWFAAQFL----- 99  
Db 1264 PSFLEKTPLGERSVN-----FSLTPNEIKVSAEAEVAPVSEVTEQVVEEHA 1311

QY 100 --LDLTLE-----SAANSLWLAIDLSERHDL-----NPVLPVETLKSDDDK 140  
Db 1312 SEDKTLVSVPSQSVTGSAGHTPYQSPTEKS-SHLPTVEIKPPAVPVSVFESDAKD 1370

QY 141 GKEREQAOKVKAFFOLVGDSEES-SILYAPVLQPLVG-EVTFDFQSAERKG----- 192  
Db 1371 ENER-----ASVSPMDEPVDSSEPIEKVLSPLRSPPLIGSESAYESFLSADDKASGRGAE 1426

QY 193 -----EISOLKSMLTTVVQER-----FA-----IQFMENAKRCVTQLDR 228  
Db 1427 SPFEKSGKQSPDQVSPVSEMTSISLYQDKQEGKSTDFAPIKEDFGQEKKTDDVEAKSS 1486

QY 229 LSAALYSTKCHSLGSGOS-----TNFGFAKSLLTRVENALVHLSGTLKAPKAETVQEV 282  
Db 1487 QPALALDE-RKLGDSVPTQIDVSQFGSFK-----EDTKMSISEGTVSDKS-ATPVDEGV 1538

QY 283 AESSVSEGLPSHMDTKHIERIPMAEQACTVSOHLHAGNLSELGNLNNMRDLAFHLIR 342  
Db 1539 AEDTVSHMGVAVSVTASVATSSFPETDSDVSPSLHA-----EVGSPHS-----T 1584

QY 343 EYSDYPRQSEPHSPISF-----LLEKAIWRGYLSLPEL-----LREMMSEQNG 385  
Db 1585 EYVDDLSVSVVQPTTFQTEKSPKEECPRMSISPPDFSPKTAKRTPVQDHRSEQS- 1643

QY 386 DALSTTFNAGLNHLDDQVLLPEVSTQVIESPQIPQAKPSYSDPRSVEHVSTQSPVDT 445  
Db 1644 -SMSIEF--GQESPQSLAMDFSR-----QSPDHPVTGAGVL-----HITENGPTVEV 1687

QY 446 QSKQDQKPOSSATSALS 462  
Db 1688 ----DYSPSDMQDSSLS 1700

RESULT 15  
US-09-134-000C-5352  
; Sequence 5352, Application US/09134000C  
; Patent No. 6617156

```
/ GENERAL INFORMATION:
/ APPLICANT: Lynn Doucette-Stamm et al
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
/ TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: 032796-032
/ CURRENT APPLICATION NUMBER: US/09/134,000C
/ CURRENT FILING DATE: 1998-08-13
/ PRIOR APPLICATION NUMBER: US 60/055,778
/ PRIOR FILING DATE: 1997-08-15
/ NUMBER OF SEQ ID NOS: 6812
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 5352
/ LENGTH: 431
/ TYPE: PRT
/ ORGANISM: Enterococcus faecalis
/ US-09-134-000C-5352

Query Match      4.9%; Score 114.5; DB 4; Length 431;
Best Local Similarity 20.2%; Pred. No. 0.031;
Matches 78; Conservative 69; Mismatches 137; Indels 103; Gaps 16;

QY 3 LSKHQIEQLSPLSDSDTCGVYKLEKSAF-----RPLRNEFNVA----- 42
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MTKRLIRKIDISDDT-----LEKEVFPQGPPEKIIISGIKPEQEKEMATLRRL 53
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 43 -----QTALRLSQ-----NPSADERDALQEAACLNKWKILSDSLYEQFSKTRDIELI 90
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 54 TYQKKETAFKQLQENQKQOETEKQALDEKNQALLETLNQKEHEIF-ELEEQLTN 112
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 91 SWFVAAQFLDPTTLESANSLWLAJSEKHWDHNLNPLVPVETLKSD-DDKGREREQADA 149
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 113 SETTEATTLQTNLQNAEKQLTVHN-----QMERLESELITQOVALEKESA 159
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 150 KVKAFFQLVGD--SEESSILYAPVLQPLVGEVTFDFQSAERKGBISQLKSM---LTTT 204
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 160 EOOHQIOTLHNLQALAEKGL-----EKKGDTGMKQVETLITTE 198
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 205 VAQERFATQFKMENAKRCVTQLDRLSALVSTKCHSLGSGQSTNFGFAKSLLTRVENALVHL 264
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 199 LEESTTTQALQEQIEKLTVQLQKVEAKQTLQEAQQEV-----IVQRMEEELKNELQN- 253
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 265 SGIKLAPKAEAKTV-----EOEVAEBSVSEGEPLS---HMDTK--HIERIPMASEQA 311
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 254 ---SLQCKNEAEVKLGRLQGVNDKHLAEKQVYEGELAAALRKEVEAKEDNLEQVQKELKTL 310
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 312 QTVSOHLHAGN--LSELGNLNNMNRDL 336
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 311 QAVSKNIEPSSEQLTELLRAKQIINDL 337
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: July 3, 2004, 02:47:19  
Job time : 26 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2004, 02:46:20 ; Search time 81 Seconds  
(without alignments)  
1779.314 Million cell updates/sec

Title: US-10-780-347-2  
Perfect score: 2343  
Sequence: 1 MFLSKHQIEQLSKPLSDSI.....DTQSKQDKPOSSATSALSW 463

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1276540 seqs, 311283816 residues

Total number of hits satisfying chosen parameters: 1276540

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:  
1: /cgn2\_5/prodata/2/pubpaa/US07\_PUBCOMB.pap:  
2: /cgn2\_5/prodata/2/pubpaa/PCT\_NEW\_PUB.pap:  
3: /cgn2\_5/prodata/2/pubpaa/US06\_NEW\_PUB.pap:  
4: /cgn2\_5/prodata/2/pubpaa/US06\_PUBCOMB.pap:  
5: /cgn2\_5/prodata/2/pubpaa/US07\_NEW\_PUB.pap:  
6: /cgn2\_5/prodata/2/pubpaa/PCTUS\_PUBCOMB.pap:  
7: /cgn2\_5/prodata/2/pubpaa/US08\_NEW\_PUB.pap:  
8: /cgn2\_5/prodata/2/pubpaa/US08\_PUBCOMB.pap:  
9: /cgn2\_5/prodata/2/pubpaa/US09\_PUBCOMB.pap:  
10: /cgn2\_5/prodata/2/pubpaa/US09\_PUBCOMB.pap:  
11: /cgn2\_5/prodata/2/pubpaa/US09\_PUBCOMB.pap:  
12: /cgn2\_5/prodata/2/pubpaa/US09\_NEW\_PUB.pap:  
13: /cgn2\_5/prodata/2/pubpaa/US10A\_PUBCOMB.pap:  
14: /cgn2\_5/prodata/2/pubpaa/US10B\_PUBCOMB.pap:  
15: /cgn2\_5/prodata/2/pubpaa/US10C\_PUBCOMB.pap:  
16: /cgn2\_5/prodata/2/pubpaa/US10\_NEW\_PUB.pap:  
17: /cgn2\_5/prodata/2/pubpaa/US60\_NEW\_PUB.pap:  
18: /cgn2\_5/prodata/2/pubpaa/US60\_PUBCOMB.pap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	143	6.1	884	14	US-10-179-766-10
2	134.5	5.7	1179	9	Sequence 10, Appl
3	134.5	5.7	1179	12	Sequence 1262, A
4	133	5.7	2295	16	Sequence 74007, A
5	129.5	5.5	860	12	Sequence 135452,
6	129.5	5.5	860	12	Sequence 838, Appl
7	129.5	5.5	860	12	Sequence 59, Appl
8	127	5.4	1294	12	Sequence 166, Appl
9	126.5	5.4	752	14	Sequence 61292, A
10	126.5	5.4	785	14	Sequence 48, Appl
11	126.5	5.4	824	15	Sequence 3569, Appl
12	126.5	5.4	1133	14	Sequence 14, Appl
13	126.5	5.4	1133	14	Sequence 16, Appl
14	126.5	5.4	1133	14	Sequence 12, Appl
15	126	5.4	506	14	Sequence 7, Appl

Sequence 1648, Appl  
Sequence 11, Appl  
Sequence 2128, Appl  
Sequence 99, Appl  
Sequence 1175, Appl  
Sequence 16, Appl  
Sequence 22, Appl  
Sequence 18, Appl  
Sequence 6346, Appl  
Sequence 254, Appl  
Sequence 48350, A  
Sequence 113490,  
Sequence 13608, A  
Sequence 115622,  
Sequence 60535, A  
Sequence 2224, Appl  
Sequence 7470, Appl  
Sequence 43924, A  
Sequence 419, Appl  
Sequence 46243, A  
Sequence 156313, A  
Sequence 22080, A  
Sequence 124349,  
Sequence 103451,  
Sequence 45920, A  
Sequence 177644,  
Sequence 1635, Appl  
Sequence 16, Appl

US-10-369-493-1648  
US-09-952-267-11  
US-10-369-493-2128  
US-09-738-630-99  
US-10-408-765A-1175  
US-10-336-472-16  
US-10-336-472-18  
US-10-336-472-22  
US-10-369-493-6346  
US-10-408-765A-234  
US-10-408-765A-2096  
US-10-282-122A-48350  
US-10-437-963-113490  
US-09-815-242-13608  
US-10-437-963-115622  
US-10-282-122A-60535  
US-10-369-493-2224  
US-10-032-585-7470  
US-10-282-122A-43924  
US-10-305-823-419  
US-10-282-122A-46243  
US-10-172-502-4  
US-10-437-963-16313  
US-10-369-493-22080  
US-10-437-963-124349  
US-10-437-963-103451  
US-10-282-122A-45920  
US-10-437-963-177644  
US-10-408-765A-1635  
US-09-820-843A-16

## ALIGNMENTS

RESULT 1  
US-10-179-766-10  
; Sequence 10, Application US/10179766  
; Publication No. US20030190312A1  
; GENERAL INFORMATION:  
; APPLICANT: Kenyon, Cynthia  
; APPLICANT: Apfeld, Javier  
; APPLICANT: Dillin, Andrew  
; APPLICANT: Garigan, Delia  
; APPLICANT: Hsu, Ao-Lin A.  
; APPLICANT: Lehrer-Graiwer, Josh  
; APPLICANT: Murphy, Coleen  
; APPLICANT: The Regents of the University of California  
; TITLE OF INVENTION: Eukaryotic Genes Involved in Adult Lifespan Regulation  
; FILE REFERENCE: 023070-119300S  
; CURRENT APPLICATION NUMBER: US/10/179,766  
; CURRENT FILING DATE: 2002-06-24  
; PRIOR APPLICATION NUMBER: US 60/300,577  
; PRIOR FILING DATE: 2001-06-22  
; PRIOR APPLICATION NUMBER: US 60/301,052  
; PRIOR FILING DATE: 2001-06-25  
; PRIOR APPLICATION NUMBER: US 60/373,975  
; PRIOR FILING DATE: 2002-04-18  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 10  
; LENGTH: 884  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
; FEATURE:  
; OTHER INFORMATION: 11w-4 (long-lived worm protein), gene F45H10.4  
US-10-179-766-10

Query Match 6.1%; Score 143; DB 14; Length 884;

Best Local Similarity 20.7%; Pred. No. 0.0021;

Matches 103; Conservative 99; Mismatches 165; Indels 130; Gaps 25;

QY 27 LEKSAPRLRNFNVAQTRKLSQNPASDERDALQEAQKLVK---KILSDSLYEQFSK 82





```
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 74007
; LENGTH: 1179
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-282-122A-74007

Query Match          5.7%; Score 134.5; DB 12; Length 1179;
Best Local Similarity 21.6%; Pred. No. 0.019;
Matches 100; Conservative 80; Mismatches 178; Indels 105; Gaps 24;

QY      5 KHOIEQLSKPLSDSDSICGVYKLEKSAFRLRNEFNVAQTALFKLSQNPASDERDALQEA 64
Db      674 KPELEQLKEIADE---ASLSEBAALKTLQDMALTELEAIK---SQGEQARIQEQ 727
QY      65 CLNKWKILSDLSYEQSKTRDIELISWFWVAQFLDITLESAAANSLEWLADLSEKHWDH 124
Db      728 GLS-----LAYCQTSQVVELETL-WKLOEE-EIDRLSEG-----DWQAD-KERCQBS 772
QY      125 L-----NPVLPVETLKSDDKGKEREQADAKVAFQVLGDSSESSILYAPVLQPL 176
Db      773 LATIASKQNLAEIEBIEKNKNAIQRYO-----NLOEVAQARLLKTEL 818
QY      177 VGEVTFDFQSAARKGBISOLKSMITTTVAQERF--AIQPKMENAKRCVTQDLRLSALVS 234
Db      819 QGQKR-YEVADIERLG---XELDNLNIQEBEIQRMLOEKVDNLEKVDTE-----LLS 866
QY      235 TKCHLSQSTNP--GFAKSL--LTRVENALVHLSG-----IKLAPKAKTV 278
Db      867 QOABESKTQTNLQGGHIRKFELDDIEGQDDIASHLDOARQONBEWIKQTRAK-- 924
QY      279 EQEVAESSVSEGELPSHMDTKHIERIPMAEQEQAQTVSOHLHAGNLS--ELGNLNNNRDL 336
Db      925 KEKVSERL-----RHQQLQTDQYOISYTEALEKAHELENLNLAEQVODLEKAIKRS 977
QY      337 AFHLRLRVSDYFROSEPHSPISFLKXAIRGWYLSLPELLREMSQONGA---LSTIFN 393
Db      978 GPNVLEAIDQY---EEVHNRLDFL--NSQRDDILSAKNLLELTITENMDVEXERFKSTFE 1032
QY      394 AA-----GLNHLDOVLLPEVSTPTVGIESPOTQAK 424
Db      1033 AIRESEKVTFKQMFQGGQAD-LILTEGDLITAGVEISVQPPGK 1074
```

## RESULT 4

```
US-10-437-963-135452
; Sequence 135452, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
```

## RESULT 5

```
US-10-072-012-838
; Sequence 838, Application US/10072012
; Publication No. US2004003493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Patturajan, Meera
; APPLICANT: Shinkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangolli, Esha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Colman, Steven D.
```

```
; SEQ ID NO 135452
; LENGTH: 2295
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_37127C.1.pep
US-10-437-963-135452
```

Query Match 5.7%; Score 133; DB 16; Length 2295;

Best Local Similarity 20.7%; Pred. No. 0.07; 176; Indels 208; Gaps 25;

Matches 118; Conservative 68; Mismatches 176; Indels 208; Gaps 25;

QY 4 SKHOIEQLSKPLSDSDSICGVYKLEKSAFRLRNEFNVAQTA-----LRKLSQNPASDER 58

Db 1417 AQKEIEELDEMQR-----NVASIAEHEDIRLEENVGAKES 1454

QY 59 DALQEAACLN--KWKIL-----SDSLYEQFSKTRDIELISWFWAA 96

Db 1455 LLLTEREQNAATLKLLAEHLEIDELIRKLEDSRKSDSLQSTIKRLEED-----GIAK 1508

QY 97 QFLDITLESAAANSLEWLADLSEKHWDHNPVLPVETLKSDDDKG-----HDDKHILELQFTIOELENT 1561

Db 1509 EALLTEKQAEHATRLTLEALEKNEELLKI-----HDDKHILELQFTIOELENT 1561

QY 142 -----KEREQADAKVAFQVLGDSSESSILYAPVLQPLVGEVTFDFP----- 185

Db 1562 AAKENLLLREREQNDATTKA--QIESQERNEQLL-----KRFVDVDRKIDLLQ 1607

QY 186 QSAERKGBISOLKSMITTTVAQERFAIOFKMENAKRCVTQDLRLSALVSTKCHSLGSOST 245

Db 1608 DTIERIGENSTIKDALLSERQEKDAIK-----KELVEAGERNEELI-----WKIEDT 1655

QY 246 NFGPAKSLLTRVENALVHLSG-----IKL-APKAEAKTVQEVAES-----SVSE 289

Db 1656 D-----KXIEHLQNALIKLEGDIEAKDISLEAREENDTIRKSLAEAKQKNEELLRKISD 1710

QY 290 GELPSHM--DTKH-----IERIPM-----ASEQAQVSOHLHAGNLSLGNLNM 332

Db 1711 NEYRIHLQDTAQKLQVDAISRLSFFVMEKQESDAKALTEARENEEDLLKRNEDLLKR 1770

QY 333 NEDLAFHL-----LREVSDFPROSEPHSPISFLLEKAIKRMGY 369

Db 1771 NODLTKIESSKTTIQETLQRLLEGKSTNLAEQVLRQOATATPSTAKSSASR--- 1827

QY 370 LSLPELLREMSQONGDALSTIFNAGLNHLDOVLLPEVSTPTVGIESPOTQAKPSVD 429

Db 1828 ----SKITRIHRSPENGHILNGDTRQAEIK-----PSTG-----TSETIPSGN 1867

QY 430 PRSV--EEHVSO---TSPVDTQSKQDQKPQ 454

Db 1868 PPDLANNEKHVEQGEKQLKVLNOKYOSQQPQ 1897



US-10-037-417-59

```
Query Match          5.5%; Score 129.5; DB 12; Length 860;
Best Local Similarity 18.5%; Pred. No. 0.032;
Matches 98; Conservative 87; Mismatches 208; Indels 137; Gaps 17;

QY 7 QIEQLSKPLSD-----DSICGVYLKLEKSAFRLNENFVNAQTALR 47
DB 139 QIEQLQKQAKAEKESQAEVDDLLAQLDLSITKAKLNAEKKA-KQLESQSELQVKLD 197
QY 48 KLSQNSPADERDALQEAELNKKWILSDSLYEQFSKTTTRDIE-----LISWFWAAQFLDIT 103
DB 198 ELQEQ-----LNDLTSQKSRLOQENSQDLTQLEAEAAQVSNLSKLSQLESQ 244
QY 104 LESANSLEWLDLSEKHWDHNLNPLVETLKSDDDKGEREQADAKVKAFFQLVGDSEE 163
DB 245 LEEAKESLE-----EESRERANLQALQLEHDLDSLEQLEEESEAKAELE-----RQ 293
QY 164 SSILYAPVLQPLUGEVTFDFQSAERKGEISQKSLMLTTTVAQ----- 207
DB 294 LSKANAELIQWR-----SKFSEGALRAELEEELKKLNQKISELEEAANAANAKCDSLE 348
QY 208 -----ERFAIQFWMENAKRCVTOL-----DRLSALVSTKCHSLG 241
DB 349 KTKSRLOQSELDLQIELERANAASAELEKKQKNFKILAEWKRYDELQALDITAQREAR 408
QY 242 SOSTNFGFAKSLTRVENALVHLSGIKLAPKAEAKTVQEVAB--SSVSEGEPLSPHMDTK 299
DB 409 NLSTE-----LFRLLKNEELKQVEALRRENKLNQDEIHDLTDQLGEGGRNVHELEK 461
QY 300 HIERIPMASQATVQSHLHAG-NLSELG-----NLNMNRDLAFHLLREVSDYFQSE 352
DB 462 ARRLAEAKDELQALAEAEAEAELEESKVLRAQVELSQIRSEIRRLAEKEEFPENTRK 521
QY 353 PHSPISPLEKATRWGVLSPLELLREMMSPONGDALSTTFNAAGLNHLDQVLLPEVSTPT 412
DB 522 NH-----QRAIE-----SLQATLEAETKGAEA-----SRLLKKLEGDINELE 559
QY 413 VGIESPOTPAKSVSDPRVBEHVQSOTSPVDTQSKQDQKQPSASATLS 462
DB 560 IALDHAN-----KANAEAKQNVKYYQQVQKELQTVQVEEQAREDAREQLA 605
```

## RESULT 7

```
US-10-080-334-166
Sequence 166, Application US/10080334
Publication No. US20040002584A1
GENERAL INFORMATION:
APPLICANT: Shimkets, Richard A
APPLICANT: Li, Li
APPLICANT: Shenoy, Suresh G
APPLICANT: Kekuda, Ramesh
APPLICANT: Spytek, Kimberly A
APPLICANT: Vernet, Corine A. M.
APPLICANT: Malyankar, Uriel M
APPLICANT: Guo, Xiaolia
APPLICANT: Gusev, Vladimir Y
APPLICANT: Casman, Stacie J
APPLICANT: Boldog, Ferenc L
APPLICANT: Furtak, Katarzyna
APPLICANT: Tchernev, Velizar T
APPLICANT: Patturajan, Meera
APPLICANT: Gangolli, Esha A
APPLICANT: Padigaru, Muralidhara
APPLICANT: Liu, Xiaohong
APPLICANT: Baumgartner, Jason C
APPLICANT: Gerlach, Valerie
APPLICANT: Spaderna, Steven K
APPLICANT: Zerhusen, Bryan D
TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
TITLE OF INVENTION: Using the Same
FILE REFERENCE: 21402-275

Query Match          5.5%; Score 129.5; DB 15; Length 860;
Best Local Similarity 18.5%; Pred. No. 0.032;
Matches 98; Conservative 87; Mismatches 208; Indels 137; Gaps 17;

QY 7 QIEQLSKPLSD-----DSICGVYLKLEKSAFRLNENFVNAQTALR 47
DB 139 QIEQLQKQAKAEKESQAEVDDLLAQLDLSITKAKLNAEKKA-KQLESQSELQVKLD 197
QY 48 KLSQNSPADERDALQEAELNKKWILSDSLYEQFSKTTTRDIE-----LISWFWAAQFLDIT 103
DB 198 ELQEQ-----LNDLTSQKSRLOQENSQDLTQLEAEAAQVSNLSKLSQLESQ 244
QY 104 LESANSLEWLDLSEKHWDHNLNPLVETLKSDDDKGEREQADAKVKAFFQLVGDSEE 163
DB 245 LEEAKESLE-----EESRERANLQALQLEHDLDSLEQLEEESEAKAELE-----RQ 293
QY 164 SSILYAPVLQPLUGEVTFDFQSAERKGEISQKSLMLTTTVAQ----- 207
DB 294 LSKANAELIQWR-----SKFSEGALRAELEEELKKLNQKISELEEAANAANAKCDSLE 348
QY 208 -----ERFAIQFWMENAKRCVTOL-----DRLSALVSTKCHSLG 241
DB 349 KTKSRLOQSELDLQIELERANAASAELEKKQKNFKILAEWKRYDELQALDITAQREAR 408
QY 242 SOSTNFGFAKSLTRVENALVHLSGIKLAPKAEAKTVQEVAB--SSVSEGEPLSPHMDTK 299
DB 409 NLSTE-----LFRLLKNEELKQVEALRRENKLNQDEIHDLTDQLGEGGRNVHELEK 461
QY 300 HIERIPMASQATVQSHLHAG-NLSELG-----NLNMNRDLAFHLLREVSDYFQSE 352
DB 462 ARRLAEAKDELQALAEAEAEAELEESKVLRAQVELSQIRSEIRRLAEKEEFPENTRK 521
```

QY 353 PHSPIFLEKAIKRWGLSLPELLREMMSEONGDALSTIFNAAGLNHLDOVLLREVSTPT 412  
 Db 522 NH-----ORALE-----SLQNTLEATKGRAB-----SRKKLEBDINELE 559  
 QY 413 VGIESPOTPOAKPSVSDRSVEHVSTPVDTSQKQKQKQSSATSALS 462  
 Db 560 IALDHAN----KANAEAKGVKKYQQVKELOTQVEEQAREADAREOLA 605

## RESULT 8

US-10-282-122A-61292  
 ; Sequence 61292, Application US/10282122A  
 ; Publication No. US20040029129A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, Liangsu  
 ; APPLICANT: Zamudio, Carlos  
 ; APPLICANT: Malone, Cheryl  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari  
 ; APPLICANT: Zyskind, Judith  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John  
 ; APPLICANT: Carr, Grant  
 ; APPLICANT: Yamamoto, Robert  
 ; APPLICANT: Forsyth, R.  
 ; APPLICANT: Xu, H.  
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A  
 ; CURRENT APPLICATION NUMBER: US/10/282,122A  
 ; CURRENT FILING DATE: 2003-02-20  
 ; PRIOR APPLICATION NUMBER: 60/191,078  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/230,335  
 ; PRIOR FILING DATE: 2000-09-06  
 ; PRIOR APPLICATION NUMBER: 60/230,347  
 ; PRIOR FILING DATE: 2000-09-09  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/267,636  
 ; PRIOR FILING DATE: 2001-02-09  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 78614  
 ; SOFTWARE: Patent in version 3.1  
 ; SEQ ID NO 61292  
 ; LENGTH: 1294  
 ; TYPE: PRT  
 ; ORGANISM: Legionella pneumophila  
 US-10-282-122A-61292

Query Match 5.4%; Score 127; DB 12; Length 1294;  
 Best Local Similarity 23.1%; Pred. No. 0.1;  
 Matches 104; Conservative 66; Mismatches 192; Indels 88; Gaps 20;  
 QY 7 QIEQLSKPLSDSICGVYK-----LEKSAFRLPNEFNVAOTLRKLS--QNP 53  
 Db 781 QVQLEBQLEAEKENALKADKQSQHEKSLDKSAIESLTSNLQKLEKQKQETLQ 840  
 QY 54 SABERDALQACLNKWLSDSLVEQFSKTRTRIELISFVAQAFLDITLLESAANSLW 113  
 Db 841 LKSLRQIOQCTL--VVEGLKEELQKQKSNTHOETIERTITREKSLADSALESRLKEMEY 899  
 QY 114 LADLSE-----KHWDLNPLPVEFTL-----KSDDDKGREREQAD-AKYKAPFQLVG 159

Db 900 LTRKKEENQKLTQVHLSLEQLEKQLOIRFEFEQLEKEKRVQSEKGSASAKRTVAS 959  
 QY 160 DEESSILYAPVLQPLVGEVTFDFQSAERKEISQLKSMLTITTTVAQERFAIQFMENA 219  
 Db 960 LREOVSNL---KLQQLQGLGEVI---QEKEG-----SSLISQSKQIIALQETIIEQ 1005  
 QY 220 KRCVTOLD-RLSALVSTKCHSIGSOST-----NFGFAKSLLTRVENALVHLSIKLA 270  
 Db 1006 KQQLBELKIKIQELYSAN-QELGKQNSLSKLNHNKVTEDLKKKLNELNVLQSLQHS 1064  
 QY 271 PKAEAKTV-----EOEVAESSVSEGLPSHMDTKH-----IERIPMASQAQTVSQHLHAG 321  
 Db 1065 SNEKEQTIRKLEELIKKDSLSKQNE-EMQLAQKHLQEBIDL-----QKEIKQQQLNTN 1118  
 QY 322 NL-SLGNLNNNRDLAFHLLREVSDYFRQSEPHSPISFLLEKAIKRWGLYSIPEL----- 375  
 Db 1119 QLESIIAQSKEAEKRYQQALQOKKGIYFARMERSPI-----YLOIQQIEQKAK 1167  
 QY 376 -LREMMSEQGDALSTIFNAAGL---NHLD 401  
 Db 1168 ELEERRETEASTAATLATKLRLEIKNYLD 1197

## RESULT 9

US-10-171-311-48  
 ; Sequence 48, Application US/10171311  
 ; Publication No. US20030087270A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Schlegel, Robert  
 ; APPLICANT: Chen, Yan  
 ; APPLICANT: Zhao, Xumei  
 ; APPLICANT: Monahan, John  
 ; APPLICANT: Kamatkar, Shubhangi  
 ; APPLICANT: Glatt, Karen  
 ; APPLICANT: Gannavarapu, Manjula  
 ; APPLICANT: Hoersh, Sebastian  
 ; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
 ; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY  
 ; TITLE OF INVENTION: OF CERVICAL CANCER  
 ; FILE REFERENCE: MRI-035  
 ; CURRENT APPLICATION NUMBER: US/10/171,311  
 ; CURRENT FILING DATE: 2002-06-12  
 ; PRIOR APPLICATION NUMBER: US 60/298,159  
 ; PRIOR FILING DATE: 2001-06-13  
 ; PRIOR APPLICATION NUMBER: US 60/298,155  
 ; PRIOR FILING DATE: 2001-06-13  
 ; PRIOR APPLICATION NUMBER: US 60/335,936  
 ; PRIOR FILING DATE: 2001-11-14  
 ; NUMBER OF SEQ ID NOS: 238  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 48  
 ; LENGTH: 752  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-171-311-48

Query Match 5.4%; Score 126.5; DB 14; Length 752;  
 Best Local Similarity 21.9%; Pred. No. 0.049;  
 Matches 111; Conservative 77; Mismatches 197; Indels 123; Gaps 24;  
 QY 5 KQIEQLSKPLS-----DDSI-----CGVYKLEKSAFRLPNEFNVAQTLRK 48  
 Db 160 KLEVEKLSKRIEMALEKLEDAFNKSKQCYSLKCN--LEKERWTTKQLSQLESKVRKE 217  
 QY 49 LSONPSADERP--ALQEAACLNKWLKLS-----DSLVEQFSKTRTRIELISFVAQFL 99  
 Db 218 LEAESLKEKTEFTFKED-JTKLTLTVMFVDERKTMSEKTKTEDKLQASQLOVEQN 276  
 QY 100 LDTT-----LESAANSLWLADEKHWDLNPLPVEFTLKSDDDKGREREQADAKYKAF 154  
 Db 277 KVTIVTEKLIETKRALKSKTDVEEKY-----SVTKERDDLKKNLKA- 319

QY 155 FOLVGDSESSILYAPVLQPLVGEVTFDFQSAER---KGEISQLKSMLTITVAQERPA 211  
DB 320 -----EEKGNDLLSRVNM-----VSTKCHSLGQSQTN-FGPAKSLITRVENALVHL 264  
QY 212 IQFKMENAKRCVTQDLRLSAL-----KTLVEQVAESSVSEGEPLSHMDTKHIERIP--MASEQAQTVS 315  
DB 371 IKELSEVERLKLKMDKMAIEDDLMKTEDEYETLERRYANERDKAQLSKLEHVXKEL 430  
QY 265 SGIKLAPKAE-----KTLVEQVAESSVSEGEPLSHMDTKHIERIP--MASEQAQTVS 315  
DB 431 AKYKLAETKTSHEQWLFKRLQEEAKS---GHLREVDALK-EKIHEYMATEDLICH 485  
QY 316 QHLHAGNLSELGNLNNNRDLAPHLREVSDFRQSEPHSPISFELLEKAIRNGYLSLPEL 375  
DB 486 QGDHVSCKKLNQOENNRDLG---REIENLTKELEYRHFSS---KSLR-----PSL 531  
QY 376 LREWMSEQGDALSTIFNAAGLNHLDOVLLPEVSTPTVGIESPOTPAKPSVSDPRSVBE 435  
DB 532 -----NGRRISD-----PQVTSKEVQTEAVDNEPPDYKSLIP-----LER 566  
QY 436 HVSQTSFVDTQSKQDQKPOSSATSALSW 463  
DB 567 AVINGQLYEESNQDDEPNDEG-SVLSF 593

## RESULT 10

US-10-309-851-24  
; Sequence 24, Application US/10309851  
; Publication No. US20030108554A1  
; GENERAL INFORMATION:  
; APPLICANT: Saus, Juan  
; APPLICANT: Revert-Ros, Francisco  
; TITLE OF INVENTION: GIPs, a Family of Polypeptides with Transcription Factor Activity  
; TITLE OF INVENTION: Interact with Goodpasture Antigen Binding Protein  
; FILE REFERENCE: 98,723-P-US  
; CURRENT APPLICATION NUMBER: US/10/309,851  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 24  
; LENGTH: 785  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-309-851-24

Query Match 5.4%; Score 126.5; DB 14; Length 785;  
Best Local Similarity 21.9%; Pred. No. 0.052;  
Matches 111; Conservative 77; Mismatches 197; Indels 123; Gaps 24;

QY 5 KHQIEQLSKPLS-----DDSI-----CGVYLKLEKSAFRLRNEFNVAQTALRK 48  
DB 64 KLEVEKLSKRMALKEKLEDAFNKSKQECYSKCN--LEKERTTKQLSQELESKVRKE 121  
QY 49 LSQNPASDERD--ALQEAACLNKWKLS-----DSLYEQFSKTRTRDIELISWFAAQFL 99  
DB 122 LEAIESRLKTEFTLKED-LTKLTLTMFVDERKTMSEKLTEDKLQAASSQLQVEQN 180  
QY 100 LDTT-----LESAANSLEWLADLSEKHWDHNLNPLVPVETLKSDDDKGKEREQADAKVAF 154  
DB 181 KVTVTTEKLIBETKRALKSKTDVEEKY-----SVTKERDDLKNKLKA- 223  
QY 155 FOLVGDSESSILYAPVLQPLVGEVTFDFQSAER---KGEISQLKSMLTITVAQERPA 211  
DB 224 -----EEKGNDLLSRVNM-----KNRLQSLAEIAEKDFLNKLNQDSGKSTTALHOENNK 274  
QY 212 IQFKMENAKRCVTQDLRLSAL-----VSTKCHSLGQSQTN-FGPAKSLITRVENALVHL 264  
DB 275 IKELSEVERLKLKMDKMAIEDDLMKTEDEYETLERRYANERDKAQLSKLEHVXKEL 334  
QY 265 SGIKLAPKAE-----KTLVEQVAESSVSEGEPLSHMDTKHIERIP--MASEQAQTVS 315  
DB 335 AKYKLAETKTSHEQWLFKRLQEEAKS---GHLREVDALK-EKIHEYMATEDLICH 389

QY 316 QHLHAGNLSELGNLNNNRDLAPHLREVSDFRQSEPHSPISFELLEKAIRNGYLSLPEL 375  
DB 390 QGDHVSQKLNQOENNRDLG---REIENLTKELEYRHFSS---KSLR-----PSL 435  
QY 376 LREWMSEQGDALSTIFNAAGLNHLDOVLLPEVSTPTVGIESPOTPAKPSVSDPRSVBE 435  
DB 436 -----NGRRISD-----PQVTSKEVQTEAVDNEPPDYKSLIP-----LER 470  
QY 436 HVSQTSFVDTQSKQDQKPOSSATSALSW 463  
DB 471 AVINGQLYEESNQDDEPNDEG-SVLSF 497

## RESULT 11

US-10-108-260A-3569  
; Sequence 3569, Application US/10108260A  
; Publication No. US20040005560A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: NO. US20040005560A1el full length cDNA  
; FILE REFERENCE: H1-A0106  
; CURRENT APPLICATION NUMBER: US/10/108,260A  
; CURRENT FILING DATE: 2002-03-27  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3569  
; LENGTH: 824  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-108-260A-3569

Query Match 5.4%; Score 126.5; DB 15; Length 824;  
Best Local Similarity 21.9%; Pred. No. 0.056;  
Matches 111; Conservative 77; Mismatches 197; Indels 123; Gaps 24;

QY 5 KHQIEQLSKPLS-----DDSI-----CGVYLKLEKSAFRLRNEFNVAQTALRK 48  
DB 105 KLEVEKLSKRMALKEKLEDAFNKSKQECYSKCN--LEKERTTKQLSQELESKVRKE 162  
QY 49 LSQNPASDERD--ALQEAACLNKWKLS-----DSLYEQFSKTRTRDIELISWFAAQFL 99  
DB 163 LEAIESRLKTEFTLKED-LTKLTLTMFVDERKTMSEKLTEDKLQAASSQLQVEQN 221  
QY 100 LDTT-----LESAANSLEWLADLSEKHWDHNLNPLVPVETLKSDDDKGKEREQADAKVAF 154  
DB 222 KVTVTTEKLIBETKRALKSKTDVEEKY-----SVTKERDDLKNKLKA- 264  
QY 155 FOLVGDSESSILYAPVLQPLVGEVTFDFQSAER---KGEISQLKSMLTITVAQERPA 211  
DB 265 -----EEKGNDLLSRVNM-----KNRLQSLAEIAEKDFLNKLNQDSGKSTTALHOENNK 315  
QY 212 IQFKMENAKRCVTQDLRLSAL-----VSTKCHSLGQSQTN-FGPAKSLITRVENALVHL 264  
DB 316 IKELSEVERLKLKMDKMAIEDDLMKTEDEYETLERRYANERDKAQLSKLEHVXKEL 375  
QY 265 SGIKLAPKAE-----KTLVEQVAESSVSEGEPLSHMDTKHIERIP--MASEQAQTVS 315  
DB 376 AKYKLAETKTSHEQWLFKRLQEEAKS---GHLREVDALK-EKIHEYMATEDLICH 430  
QY 316 QHLHAGNLSELGNLNNNRDLAPHLREVSDFRQSEPHSPISFELLEKAIRNGYLSLPEL 375  
DB 431 QGDHVSQKLNQOENNRDLG---REIENLTKELEYRHFSS---KSLR-----PSL 476  
QY 376 LREWMSEQGDALSTIFNAAGLNHLDOVLLPEVSTPTVGIESPOTPAKPSVSDPRSVBE 435  
DB 477 -----NGRRISD-----PQVTSKEVQTEAVDNEPPDYKSLIP-----LER 511  
QY 436 HVSQTSFVDTQSKQDQKPOSSATSALSW 463  
DB 512 AVINGQLYEESNQDDEPNDEG-SVLSF 538

## RESULT 12

```
US-10-309-851-14
; Sequence 14, Application US/10309851
; Publication No. US20030108554A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; APPLICANT: Revert-Ros, Francisco
; TITLE OF INVENTION: GIPs, a Family of Polypeptides with Transcription Factor Activity
; TITLE OF INVENTION: Interact with Goodpasture Antigen Binding Protein
; FILE REFERENCE: 98,723-F-US
; CURRENT FILING DATE: 2002-12-04
; CURRENT APPLICATION NUMBER: US/10/309,851
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 1133
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-309-851-14

Query Match          5.4%; Score 126.5; DB 14; Length 1133;
Best Local Similarity 21.9%; Pred. No. 0.091;
Matches 111; Conservative 77; Mismatches 197; Indels 123; Gaps 24;

QY 5 KHQIEQLSKPLS-----DDSI-----CGVYLKLEKSAFRLPNEFNVAQTALRK 48
DB 414 KLEVEKLSKRMALKELEDAFNKSKQECYSLKCN--LEKERMTTKQLSQELESLEKVRKE 471
QY 49 LSONPSADERD--ALQEAACLNKWKILS-----DSLYEQFSKTRTRDIELISWFWAAQFL 99
DB 472 LEAIESRLEKTEFTLKED-LTKLKTLTVMFVDERKTMSEKLEKTKEDKLAQASSQLQVEQN 530
QY 100 LDTT-----LSAANSLEWLADLSEKHWDLNPNVLPVETLKSDDDKGKEREQADAKVAF 154
DB 531 KVTITVEKLIBETKALKSKTDVEEKY-----SVTKERDDLKNKLKA- 573
QY 155 FQVGDSESSILYAPVLQPLVGEVTFDFQSAER---KGEISQLKSMLTITVAQERPA 211
DB 574 -----EEKGNDLLSRVNL-----KNRLQSLEALEKDFLNKNLQDQSGKSTTALHOENNK 624
QY 212 IQFKMENAKRCVTOLDRLSAL-----VSTKCHSLGSGSTN-FGPAKSLTRVENALVHL 264
DB 625 IKELSQEVEVERLKLKDKMAIEDDLTKTEDEYETLERRYANERDKAQFLSKELHVKMEL 684
QY 265 SGIKLAPKAEA-----KTVEQVVAESSVSEGEPLPSHMDTKHIERIP--MASEQAQTVS 315
DB 685 AKYKLAETKETSHEQWLFKRLQEEBAKS-----GHLSREVDALK-EKIHETMATEDLICH 739
QY 316 QHLHAGNLSELGNLNNNRDLAFHLLREVSDYFRQSEPHSPISFLLEKAIKRWGYSLSPEL 375
DB 625 IKELSQEVEVERLKLKDKMAIEDDLTKTEDEYETLERRYANERDKAQFLSKELHVKMEL 684
QY 265 SGIKLAPKAEA-----KTVEQVVAESSVSEGEPLPSHMDTKHIERIP--MASEQAQTVS 315
DB 685 AKYKLAETKETSHEQWLFKRLQEEBAKS-----GHLSREVDALK-EKIHETMATEDLICH 739
QY 316 QHLHAGNLSELGNLNNNRDLAFHLLREVSDYFRQSEPHSPISFLLEKAIKRWGYSLSPEL 375
DB 740 QGDHSLVQKLNQENNRDLG---REIENLTKELERYRHS-----KSLR-----PSL 785
QY 376 LREMMSQNGDALSTIFNAAGNLHLDQVLLPEVSTPTVGIESPQTPOAKPSVSDPRSVEE 435
DB 786 -----NGRRISD-----PQVFSKEVQTEAVDNEPPDYKSLIP-----LER 820
QY 436 HVSQTSPTVDTOSKQDQKQSSATSALSW 463
DB 821 AVINGQLYESENQDEPDNEG-SVLSF 847

RESULT 13
US-10-309-851-16
; Sequence 16, Application US/10309851
; Publication No. US20030108554A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; APPLICANT: Revert-Ros, Francisco
; TITLE OF INVENTION: GIPs, a Family of Polypeptides with Transcription Factor Activity
; TITLE OF INVENTION: Interact with Goodpasture Antigen Binding Protein
; FILE REFERENCE: 98,723-F-US
; CURRENT FILING DATE: 2002-12-04
; CURRENT APPLICATION NUMBER: US/10/309,851
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 1133
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-309-851-16

Query Match          5.4%; Score 126.5; DB 14; Length 1135;
Best Local Similarity 21.9%; Pred. No. 0.091;
Matches 111; Conservative 77; Mismatches 197; Indels 123; Gaps 24;

QY 5 KHQIEQLSKPLS-----DDSI-----CGVYLKLEKSAFRLPNEFNVAQTALRK 48
DB 414 KLEVEKLSKRMALKELEDAFNKSKQECYSLKCN--LEKERMTTKQLSQELESLEKVRKE 471
QY 49 LSONPSADERD--ALQEAACLNKWKILS-----DSLYEQFSKTRTRDIELISWFWAAQFL 99
DB 472 LEAIESRLEKTEFTLKED-LTKLKTLTVMFVDERKTMSEKLEKTKEDKLAQASSQLQVEQN 530
QY 100 LDTT-----LSAANSLEWLADLSEKHWDLNPNVLPVETLKSDDDKGKEREQADAKVAF 154
DB 531 KVTITVEKLIBETKALKSKTDVEEKY-----SVTKERDDLKNKLKA- 573
QY 155 FQVGDSESSILYAPVLQPLVGEVTFDFQSAER---KGEISQLKSMLTITVAQERPA 211
DB 574 -----EEKGNDLLSRVNL-----KNRLQSLEALEKDFLNKNLQDQSGKSTTALHOENNK 624
QY 212 IQFKMENAKRCVTOLDRLSAL-----VSTKCHSLGSGSTN-FGPAKSLTRVENALVHL 264
DB 625 IKELSQEVEVERLKLKDKMAIEDDLTKTEDEYETLERRYANERDKAQFLSKELHVKMEL 684
QY 265 SGIKLAPKAEA-----KTVEQVVAESSVSEGEPLPSHMDTKHIERIP--MASEQAQTVS 315
DB 685 AKYKLAETKETSHEQWLFKRLQEEBAKS-----GHLSREVDALK-EKIHETMATEDLICH 739
QY 316 QHLHAGNLSELGNLNNNRDLAFHLLREVSDYFRQSEPHSPISFLLEKAIKRWGYSLSPEL 375
DB 740 QGDHSLVQKLNQENNRDLG---REIENLTKELERYRHS-----KSLR-----PSL 785
QY 376 LREMMSQNGDALSTIFNAAGNLHLDQVLLPEVSTPTVGIESPQTPOAKPSVSDPRSVEE 435
DB 786 -----NGRRISD-----PQVFSKEVQTEAVDNEPPDYKSLIP-----LER 820
QY 436 HVSQTSPTVDTOSKQDQKQSSATSALSW 463
DB 821 AVINGQLYESENQDEPDNEG-SVLSF 847

RESULT 14
US-10-309-851-12
; Sequence 12, Application US/10309851
; Publication No. US20030108554A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; APPLICANT: Revert-Ros, Francisco
; TITLE OF INVENTION: GIPs, a Family of Polypeptides with Transcription Factor Activity
; TITLE OF INVENTION: Interact with Goodpasture Antigen Binding Protein
; FILE REFERENCE: 98,723-F-US
; CURRENT FILING DATE: 2002-12-04
; CURRENT APPLICATION NUMBER: US/10/309,851
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 1135
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-309-851-12

Query Match          5.4%; Score 126.5; DB 14; Length 1135;
Best Local Similarity 21.9%; Pred. No. 0.091;
Matches 111; Conservative 77; Mismatches 197; Indels 123; Gaps 24;

QY 5 KHQIEQLSKPLS-----DDSI-----CGVYLKLEKSAFRLPNEFNVAQTALRK 48
```

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 1133
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-309-851-16

Query Match          5.4%; Score 126.5; DB 14; Length 1133;
Best Local Similarity 21.9%; Pred. No. 0.091;
Matches 111; Conservative 77; Mismatches 197; Indels 123; Gaps 24;

QY 5 KHQIEQLSKPLS-----DDSI-----CGVYLKLEKSAFRLPNEFNVAQTALRK 48
DB 414 KLEVEKLSKRMALKELEDAFNKSKQECYSLKCN--LEKERMTTKQLSQELESLEKVRKE 471
QY 49 LSONPSADERD--ALQEAACLNKWKILS-----DSLYEQFSKTRTRDIELISWFWAAQFL 99
DB 472 LEAIESRLEKTEFTLKED-LTKLKTLTVMFVDERKTMSEKLEKTKEDKLAQASSQLQVEQN 530
QY 100 LDTT-----LSAANSLEWLADLSEKHWDLNPNVLPVETLKSDDDKGKEREQADAKVAF 154
DB 531 KVTITVEKLIBETKALKSKTDVEEKY-----SVTKERDDLKNKLKA- 573
QY 155 FQVGDSESSILYAPVLQPLVGEVTFDFQSAER---KGEISQLKSMLTITVAQERPA 211
DB 574 -----EEKGNDLLSRVNL-----KNRLQSLEALEKDFLNKNLQDQSGKSTTALHOENNK 624
QY 212 IQFKMENAKRCVTOLDRLSAL-----VSTKCHSLGSGSTN-FGPAKSLTRVENALVHL 264
DB 625 IKELSQEVEVERLKLKDKMAIEDDLTKTEDEYETLERRYANERDKAQFLSKELHVKMEL 684
QY 265 SGIKLAPKAEA-----KTVEQVVAESSVSEGEPLPSHMDTKHIERIP--MASEQAQTVS 315
DB 685 AKYKLAETKETSHEQWLFKRLQEEBAKS-----GHLSREVDALK-EKIHETMATEDLICH 739
QY 316 QHLHAGNLSELGNLNNNRDLAFHLLREVSDYFRQSEPHSPISFLLEKAIKRWGYSLSPEL 375
DB 740 QGDHSLVQKLNQENNRDLG---REIENLTKELERYRHS-----KSLR-----PSL 785
QY 376 LREMMSQNGDALSTIFNAAGNLHLDQVLLPEVSTPTVGIESPQTPOAKPSVSDPRSVEE 435
DB 786 -----NGRRISD-----PQVFSKEVQTEAVDNEPPDYKSLIP-----LER 820
QY 436 HVSQTSPTVDTOSKQDQKQSSATSALSW 463
DB 821 AVINGQLYESENQDEPDNEG-SVLSF 847

RESULT 14
US-10-309-851-12
; Sequence 12, Application US/10309851
; Publication No. US20030108554A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; APPLICANT: Revert-Ros, Francisco
; TITLE OF INVENTION: GIPs, a Family of Polypeptides with Transcription Factor Activity
; TITLE OF INVENTION: Interact with Goodpasture Antigen Binding Protein
; FILE REFERENCE: 98,723-F-US
; CURRENT FILING DATE: 2002-12-04
; CURRENT APPLICATION NUMBER: US/10/309,851
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 1135
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-309-851-12

Query Match          5.4%; Score 126.5; DB 14; Length 1135;
Best Local Similarity 21.9%; Pred. No. 0.091;
Matches 111; Conservative 77; Mismatches 197; Indels 123; Gaps 24;

QY 5 KHQIEQLSKPLS-----DDSI-----CGVYLKLEKSAFRLPNEFNVAQTALRK 48
```

Db 414 KLEVEKLSKRMALKELEDAFNKSKQBCYSLKN--LEKERTTTQLSOELESKVRKE 471  
QY 49 LSNPSADERD--ALQEAACLNKWKILS-----DSLYEQFSKTTTRDIELISFWVAAQPL 99  
Db 472 LEAIESRLEKTEFTLKED--LTKLKTITVMFVDERKTMSEKLLKTKEDKLOAASSQLQVEQN 530  
QY 100 LDTT-----LESNANSLEWLADLSEKHWDHLPVLPVETLKSDDDKGKEREQADAKVAF 154  
Db 531 KVTITVEKLIBETKRAKSKTQVDEEKY-----SVTKERDDLKNKLKA- 573  
QY 155 FOLVGDSESSILYAPVLOPLVGEVTFDFQSAER---KGEISQLKSMLTITTTVAQERPA 211  
Db 574 -----SEKGNLDSRVNML-----KNRLQSLERAIEKDFLKNLNQDSGKSTTALHGOENNK 624  
QY 212 IQFMENAKRCVTQDLRLGAL-----VSTKCHSLGQSQTN-FGPAKSLTTRVENALVHL 264  
Db 625 IKELSOEVLRLKLDKMDKAIEDLMTKTEDEYETLERRVYANERDKAFLSKSELEHYVMEL 684  
QY 265 SGIKLAPKAE-----KTVEQEAESSVSEGELPSHMDTKHIERIP--MASEQAQTVS 315  
Db 685 AKYKLAKEKETSHEQWLFKELQEBEAXS-----CHLSREVDALK-EKIHVEYMATEDLICH 739  
QY 316 QHLHAGNLSLGNLNNRDALPHLLREVSDFPROSEPHSPISFLLKKAIRWGLSLP 375  
Db 740 QGDHSLVQLKLNQOENRNLG---REIENLTKELEYRHFS---KSLR-----PSL 785  
QY 376 LREMSEQNGDALSTIFNAGLNHLQVLLPEVSTPTVGIESPOTPAKPSVSDPRSVEE 435  
Db 786 -----NRRISD-----PQVFSKEVQTEAVDNEPPDYKSLP-----LSR 820  
QY 436 HVSQTSVPDTQSKQDQKQPSATSALSW 463  
Db 821 AVINGQIYBESNQDEPDNEG-SVLSF 847

RESULT 15  
US-10-157-223-7  
; Sequence 7, Application US/10157223  
; Publication No. US20030036510A1  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; Hillman, Jennifer L.  
; Au-Yang, Janice  
; Zweiger, Gary B.  
; TITLE OF INVENTION: NOVEL HUMAN CELL DIVISION CYCLE PROTEINS  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: U.S.  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/157,223  
; FILING DATE: 28-May-2002  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/712,708  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0122 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-853-0355  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:  
LENGTH: 506 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 1077057  
SEQUENCE DESCRIPTION: SEQ ID NO: 7:  
US-10-157-223-7  
Query Match 5.4%; Score 126; DB 14; Length 506;  
Best Local Similarity 19.1%; Pred. No. 0.03; Indels 148; Gaps 17;  
Matches 90; Conservative 71; Mismatches 161;  
QY 15 LSDDSICGVLYLKLSKAPRPLRNEFNVAQTAIRKLSONPSADERDALQEAACLNKWKILSD 74  
Db 13 LSDSDVEVHNVDKSP-----INWK--QQ 36  
QY 75 SLYEQFSKTTTRDIELISFWVAAQFLDITLESANSLWADLSEKHWDHLPVLPVETL 134  
Db 37 SIHEQRFKRNQDIKNLETVQDMYSHLNKRVDR-----LSNLPSSUTDLPVTKFLNA 90  
QY 135 KSDDDKGRERQADAKVAPFQLVGDSESSILYAPVLOPLVGEVTFDFQSAERKGEI 194  
Db 91 NFDKMEKSGENVDEIATYNEMVEDLPE-----QLA-----KDLCKEKG 131  
QY 195 SOLKSMLTITTTVAQERFAIQFVENAKRCVTOLDLSALVSTKCHSLGQSQTNFGPAKSL 254  
Db 132 SKSPSLIRDAILKRAKIDSVTVEAKK-----KLDELYKEKNAHISSEDHTGFDSSFM 185  
QY 255 TRVENALVHLSGI---KLAPKAEAKTVQEAESSVSS-----GEL 292  
Db 186 NKQKGAKPLEATPSEALSSAAES--NINLKAKSSVPQTFIDFKDDPKMLAKEEFGKI 244  
QY 293 PSHMDTKH-----IERIPMASEQ-----ACTVSQHLHAG----- 321  
Db 245 STINEYSKSKQKFLLEHLPIISEQKQDALMMKAFYQLHGDDKMTLQVHQSELMAIKEYI 304  
QY 322 -----NLSELGNLNNRDALPHLLREVSDFPROSEPHSPISFLLKKAIRWGLSLP 373  
Db 305 DMKKIPYLNPMELSNVINM-----FFEKV--IFNKDKPMGKESFL--RSVQERFLHIQ 353  
QY 374 E-----LLREMSEQNGDALSTIFNAGLNHLQVLLPEVSTPTVGIESPQ 419  
Db 354 KRKSLQOEEMDESNAEGVETI-----QLKSLDDSTELEWNLDPFNKSKPE 399  
Search completed: July 3, 2004, 02:52:30  
Job time : 83 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2004, 02:16:54 ; Search time 27 seconds  
(without alignments)  
1649.507 Million cell updates/sec

Title: US-10-780-347-2  
Perfect score: 2343  
Sequence: 1 MPLSKHQIEQLSKPLSDDSI .....DTQSKQDKQSSATSLSW 463  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191536 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	241.5	10.3	351	AC3090	conserved hypotet
2	241.5	10.3	351	G95144	hypothetical prote
3	209	8.9	366	D83350	hypothetical prote
4	152	6.5	2245	T18278	myosin heavy chain
5	143	6.1	884	T20405	hypothetical prote
6	141.5	6.0	344	G83634	hypothetical prote
7	141.5	6.0	1510	T16927	hypothetical prote
8	134.5	5.7	1179	Z38012	hypothetical prote
9	134	5.7	1051	T18302	apsB protein - Eme
10	129.5	5.5	1133	T22976	hypothetical prote
11	127.5	5.4	2471	T22977	large tegument pro
12	126.5	5.4	1023	A81643	ATP-dependent dsDN
13	126	5.4	506	S50914	cell division cont
14	125.5	5.4	1287	T22235	hypothetical prote
15	125.5	5.4	1992	A47297	myosin heavy chain
16	125	5.3	1194	T37503	probable chromosom
17	124.5	5.3	550	T40379	hypothetical prote
18	124.5	5.3	1940	S04090	myosin heavy chain
19	124.5	5.3	1940	A24922	myosin heavy chain
20	124.5	5.3	1964	A59282	nonmuscle myosin I
21	123.5	5.3	867	C71638	DNA polymerase I
22	123.5	5.3	1051	T18351	impl protein - Myc
23	123	5.2	1935	A37102	myosin beta heavy
24	123	5.2	1935	S06006	myosin beta heavy
25	123	5.2	2022	T43214	ovtl protein - nem
26	122.5	5.2	2469	H36812	hypothetical prote
27	122	5.2	1934	T48153	myosin heavy chain
28	122	5.2	2109	E89066	protein H05009.1
29	122	5.2	2109	T33247	hypothetical prote

30 121.5 5.2 1179 2 G95144 conserved hypotet  
31 121.5 5.2 3225 2 I52300 giantin - human  
32 121.5 5.2 3259 1 A56539 giantin - human  
33 120.5 5.1 999 2 J05278 oxygen-regulated p  
34 120.5 5.1 1938 2 A59293 skeletal myosin he  
35 120 5.1 650 2 S44806 F10E9.6 protein -  
36 120 5.1 828 2 F96535 hypothetical prote  
37 120 5.1 1034 2 T32297 hypothetical prote  
38 119.5 5.1 1066 1 A48669 kinesin-related pr  
39 119.5 5.1 1690 2 T13030 microtubule bindin  
40 119 5.1 1023 2 A81280 ATP-dependent dsDN  
41 119 5.1 1356 2 S32763 kinesin 1 - human  
42 119 5.1 1935 2 A59286 myosin heavy chain  
43 119 5.1 2104 2 T38774 myosin-3 heavy cha  
44 118.5 5.1 1920 2 A53188 pericentrin - mous  
45 118.5 5.1 1937 2 I38055 myosin heavy chain

## ALIGNMENTS

## RESULT 1

AC3090  
conserved hypothetical protein Atu4343 [imported] - Agrobacterium tumefaciens (strain C  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
C:Accession: AC3090  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, J.  
; Karp, P.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: A82577; MUID:21608550; PMID:11743193  
A:Accession: AC3090  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-351 <KUR>  
A:Cross-references: GB:AB008689; PIDN:AAL45137.1; PID:gl7742810; GSPDB:GN00187  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: Atu4343  
A:Map position: linear chromosome

Query Match 10.3%; Score 241.5; DB 2; Length 351;  
Best Local Similarity 23.6%; Pred. No. 2.3e-08;  
Matches 96; Conservative 64; Mismatches 162; Indels 85; Gaps 15;

QY 5 KHQIEQLSKPLSDDSICGVYKLB---KSAFRPLARNEFNVAQTALRKLSONPSADERDAL 61  
DB 7 KRDIIEFLGN-----CGDNIRNDSRTREIYRIKDERNQAR-----AERAAAS 48  
QY 62 QEACL---NKWKILSDSLYEQFSKTRDIELISFWAAQFLDITLLESAANSLEWLADLS 118  
DB 49 PQDNLKISSWDSVNLGLQIYSEKDXVEILAWLAELASRL-RGFHGLREIYELCGDLF 107  
QY 119 EKHFDHLPVLPVETLTKSDDDKGRERQAKVAPFQVNDSEESILYAPVQLPLV- 177  
DB 108 YNHWDLSRSI-----SDND-----BEKFPAPFAGLNGISGEPT-LVQPLRLASLIP 152  
QY 178 ---CEVTFDFQSAERKEISQLKSMITTTVAQERPAIQFKMENAKRCVTQLDRLSALV 233  
DB 153 GKGFGHSLWDFQLAQRPNEKREELRYRTASEAGVAAMSSHLAAVNTCLSSFDATAVL 212  
QY 234 STKCHSLGSQSTNGFPAKSLITVENALVLSIGIKLAPKAEATV---EQEVAESSVSEG 290  
DB 213 SERCGQAAPPSN-----IRNTLIBAA-----AAITLGRQOEAP----- 249  
QY 291 ELPSHMDTKHIERIFPMASEQATVSOHLHAGNLSLSELNNMNRDLAFHLLRLRVSDYFRQ 350  
DB 250 -----VEQFPALTA--AGIDESQSAARTSPASPEGISSEDEAFELLSVARYFR 297

QY 351 SEPHSPISFLLEKAKRWGYSLPPELLRENMSEONGDALSTFNAAGL 397  
 Db 298 TEPHSPISLSIETLVRRGRMDFSELLAELLPET--QARNAVLTAAGI 342

## RESULT 2

G98196  
 Hypothetical protein AGR\_L1042 [imported] - Agrobacterium tumefaciens (strain C58, Cere  
 C/Species: Agrobacterium tumefaciens  
 C/Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 18-Nov-2002  
 C/Accession: G98196  
 R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ouello, B.; Goldman,  
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
 Science 294, 2323-2328, 2001  
 A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
 A/Reference number: A97359; MUID:21608551; PMID:11743194  
 A/Accession: G98196  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-351 <KUR>  
 A/Cross-references: GB:AE007870; PIDN:AAK89097.1; PID:gl5158901; GSPDB:GN00170  
 C/Genetics:  
 A/Gene: AGR\_L1042  
 A/Map position: linear chromosome

Query Match 10.3%; Score 241.5; DB 2; Length 351;  
 Best Local Similarity 23.6%; Pred. No. 2.3e-08;  
 Matches 96; Conservative 64; Mismatches 162; Indels 85; Gaps 15;

QY 5 KQIEQLSKPLSDSDICGVYKLE---KSAFRPLRNEFNVAQTALRKLSONPNSADERDAL 61

Db 7 KDIIEFLGN-----CGDNIRNDSRTREIYYIRIKDERNQAR-----AEERAA 48

QY 62 QEACL---NKWKILSDSLVEQFSKTRDIELISFNVAQFLDITLLESAANSLEWLADLS 118

Db 49 PQDNLKISSWDSVNLGLQIYSEKDVLEILAWLASRL-RGFHGLREIYELGDLF 107

QY 119 EKHMDHLPVLPVETLKSDDDKGKEREQADAKVKAFFQLVGDSSESSILYAPVLQPLV 177

Db 108 YNHWDLSRSI-----SDDND-----EKKPAPFAGLNGIGSEGT-LVQPLRLASLIP 152

QY 178 ---GEVTFDFQSAERKGEISQLKSLMTTVAQERFAIQKVENAKRCVTQDRLSALV 233

Db 153 GKGFGHSLWDFQLAQPNESKRREBYRIASEAGVAANGSHLAANVTCLSSFDATAVL 212

QY 224 STKCHSLGSGSTNFGFAKSLTRVENALVHLSGLKAPKAEATV---EQEVAESSVSEG 290

Db 213 SERCGQAAPPSSN-----IRNTLIEA-----AAIRTLGGRDQEPAP----- 249

QY 291 ELPSHMDTKHTERIPMAEQATQVSHLHAGNLSELGNLNMNRDLAFHLLREVSDFRQ 350

Db 250 ---VEQTPATA--AGTDESQQAARTSPASPEGISRDEAFETLLSVARYFR 297

QY 351 SEPHSPISFLLEKAKRWGYSLPPELLRENMSEONGDALSTFNAAGL 397

Db 298 TEPHSPISLSIETLVRRGRMDFSELLAELLPET--QARNAVLTAAGI 342

## RESULT 3

D83350  
 Hypothetical protein PA2360 [imported] - Pseudomonas aeruginosa (strain PA01)  
 C/Species: Pseudomonas aeruginosa  
 C/Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C/Accession: D83350  
 R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bu  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
 ; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
 A/Reference number: A82950; MUID:20437337; PMID:10984043  
 A/Accession: D83350  
 A/Status: preliminary  
 A/Molecule type: DNA

A/Residues: 1-366 <STO>

A/Cross-references: GB:AE004662; GB:AE004091; NID:g9948395; PIDN:AAG05748.1; GSPDB:GN001  
 A/Experimental source: strain PA01

C/Genetics:

A/Gene: PA2360

Query Match 8.9%; Score 209; DB 2; Length 366;

Best Local Similarity 21.5%; Pred. No. 3.1e-06;

Matches 91; Conservative 73; Mismatches 151; Indels 108; Gaps 16;

QY 1 MPLSKH--QIEQLSKPLSDSDICGVYKLEKSAFRPLRNEFNVAQTALRKLSONPSADER 58

Db 3 LPLSGNALSLVLELPIDPGQACGSLRDPDYDR-----LRELREDDSSLP 50

QY 59 DALQEAELNK--WKILSDSLVEQFSKTRDIELISFNVAQFLDITLLESAANSLEWLAD 116

Db 51 TGVWQAEAKRADWAAVEQLASELLQRRSKDLMAAWLGEA-WLQGGGLGQLQALVLLAE 109

QY 117 LSEKHWHDHLPVLPVETLKSDDDKGKEREQADAKVKAFFQLVGDSSESSILYAPVLQPL 176

Db 110 LCEPIPEEVP-----QAQDQSWRVPPIDWLRRVAELH-----TRUPL 151

QY 177 VG-----EVTFFDFQSAERK---GEISQLKSLMTTVAQERFAIQKVENAKR----- 221

Db 152 MCGGAFAEITLYAWQLRQVQVSGDSKSAKAAEAALQOK-----KLDEALRAEPLVQ 206

QY 222 -----CVTOL-----DRLSALVSTKCHSLGSGSTNFGFAKSLTRVENALVHL 264

Db 207 WORKQASLLACQQLRLLEQWCDRLGELAPSCPLREV-----IAQWALLKEFIAMH- 260

QY 265 SGIKLAPKAEAKTVQEVAESSVSEGLPFSHMDTKHTIRIPWASQOATVSOHLHAGNLS 324

Db 261 -----PQAPLSEEQPPVAEADASEG-----DITGEESVPASAPSGPAGAP----- 300

QY 325 ELGNLNMNRDLAFHLLREVSDFRQSEPHSPISFLLEKAKRWGYSLPPELLRENMSEON 384

Db 301 -----TSREDAYRQLLIADYLRTPHSPFVYLIKRAVEWGNKPLSELAEI---N 350

QY 385 GDA 387

Db 351 ADS 353

## RESULT 4

T18278  
 myosin heavy chain [similarity] - slime mold (Dictyostelium discoideum)  
 C/Species: Dictyostelium discoideum  
 C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 08-Sep-2000  
 C/Accession: T18278; T30579  
 R/Hammer III, J.A.; Jung, G.  
 J. Biol. Chem. 271, 7120-7127, 1996  
 A/Title: The sequence of the dictyostelium myo J heavy chain gene predicts a novel, dime  
 A/Reference number: Z18854; MUID:96215148; PMID:8636147  
 A/Status: preliminary;  
 A/Molecule type: DNA  
 A/Residues: 1-2245 <HAM>  
 A/Cross-references: EMBL:U42409; NID:gl150765; PID:gl150766; PIDN:AAA85186.1  
 R/Titus, M.A.; Kuspa, A.; Loomis, W.F.  
 Proc. Natl. Acad. Sci. U.S.A. 91, 9446-9450, 1994  
 A/Title: Discovery of myosin genes by physical mapping in Dictyostelium.  
 A/Reference number: Z20873; MUID:95023928; PMID:7937787  
 A/Accession: T30579  
 A/Status: preliminary; translated from GB/EMBL/DBD  
 A/Molecule type: DNA  
 A/Residues: 1-2245 <HAM>

A/Cross-references: EMBL:U42409; NID:gl150765; PID:gl150766; PIDN:AAA85186.1  
 R/Titus, M.A.; Kuspa, A.; Loomis, W.F.  
 Proc. Natl. Acad. Sci. U.S.A. 91, 9446-9450, 1994  
 A/Title: Discovery of myosin genes by physical mapping in Dictyostelium.  
 A/Reference number: Z20873; MUID:95023928; PMID:7937787  
 A/Accession: T30579  
 A/Status: preliminary; translated from GB/EMBL/DBD  
 A/Molecule type: DNA  
 A/Residues: 1-190,'F',192-283,'T',285-290,'R',292-331,'IE','WW',337-338,'LK',342,'YRMS'  
 A/Cross-references: EMBL:L35322; NID:gl039360; PID:gl039361; PIDN:AAA79858.1  
 C/Genetics:  
 A/Gene: myoJ  
 A/Introns: 257/3; 307/3  
 C/Superfamily: myosin heavy chain; myosin motor domain homology  
 C/Keywords: ATP; P-loop  
 F:84-809/Domain: myosin motor domain homology <MWO>

Query Match 6.5%; Score 152; DB 2; Length 2245;  
Best Local Similarity 23.6%; Pred. No. 0.21;  
Matches 81; Conservative 57; Mismatches 143; Indels 62; Gaps 13;

QY 23 VYKLEKSAFRPLRNFVNAQTALRKLSONPQADDERALQEAACLNKWKILSDSLYEQFSK 82  
DB 964 IQURARSLRTVQEGOKNKLQKLELQWLRTSEAKRQK---LQDKVKSPTTISELSS 1020  
QY 83 TTRDIEL-LSWFVAQFLDITLESANSL-ENLADLSEK-----HWDHLNPLVPETIK 135  
DB 1021 NNDHLELQLEIQKYLQYELDNSQSQQLSECLSKLEBQTOQLDHSKLNKLEKLDSD 1080  
QY 136 SDDDKKEREQADAKYKAPQLVGDSESSILYAPVLQPLVGEVTF---PDFOSAER- 190  
DB 1081 QHDSIEKLOSQFNETHQQLQFQKQSEELSSKLSKTTQ-----QLDNKQEFDRLSQERD 1135  
QY 191 -----KGESISOLKSLMTTVAERFAIOFKMENAKRCVTQDLRSALYSTKCHSLGQS 244  
DB 1136 TDMNNQLTQQLKK-ANSTLEDEYFSLGIRDNLERQVLELRDENQLIKERLDSLGQS 1194  
QY 245 TNF--GFA-----KSLLTRVENALVHLSGIKLAPKAEAK-----TVE 279  
DB 1195 SQFQGALEKQQLQVQSQQLIKLSEKLGSEERAKQINQLELTDHKSKLQIQ 1254  
QY 280 QEVAESSVS-----EGELPSHMDTK-----HIERIPWASEQAAQ 312  
DB 1255 LQTEQSNKIKKLEKLEYQDEKKQLQQLERIKQSKQSV 1297

## RESULT 5

T20405  
Hypothetical protein E01G4.6 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T20405; T22248  
Submitted to the EMBL Data Library, December 1996  
R:Sims, M.  
A:Reference number: Z19569  
A:Accession: T20405  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-884 <W1>  
A:Cross-references: EMBL;Z81223; PIDN:CAB05719.1; GSPDB:GN00020; CESP:E01G4.6  
A:Experimental source: clone E01G4  
R:Sims, M.  
Submitted to the EMBL Data Library, November 1996  
A:Reference number: Z19536  
A:Accession: T22248  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-884 <W12>  
A:Cross-references: EMBL;Z81538; PIDN:CAB04387.1; GSPDB:GN00020; CESP:E01G4.6  
A:Experimental source: clone F45H10  
C:Genetics:  
A:Gene: CESP:E01G4.6  
A:Map position: 2  
A:Introns: 24/3; 77/1; 745/3

Query Match 6.1%; Score 143; DB 2; Length 884;

Best Local Similarity 20.7%; Pred. No. 0.21;  
Matches 103; Conservative 99; Mismatches 165; Indels 130; Gaps 25;

QY 27 LEKSAFRPLRNFVNAQTALRKLSONPQADDERALQEAACLNKWKILSDSLYEQFSK 82  
DB 218 LRPDAYTEARITRRLALTSTLSOR-----RERFLKYSQIRQVVAQNLFQGOOR 268  
QY 83 TTRDIELLSWFVAQFLDITLESANSL-ENLADLSEKHLNPLVPETIKSDDKGK 142  
DB 269 LTRPVEDAVTSTSSKLAVEEAPETTQEEITTDASE-----VITTKAVEATE 317  
QY 143 E-REQADAKVKAFFQVGDSESSILYAPVLQ-----LPLVGEVTFD-FQ 186

DB 318 EVTEATEATEA---PVATTKESEMHWNTIRNIRSASEKDLKYYVTLISEGKFSLEFE 374  
QY 187 SAERK-----GASISOLKSLMTTVAQ-----ERFAIOFKMENAKRCVTQQL 226  
DB 375 LAEQKULTLKSPDEKLSKMAKULADLINEALSSEKSGEIQAMEKFEPEKSELVAME 434  
QY 227 DRLGALYSTKCHSLGQSQTNFGFAKSLLTR-----VENALVHLSGIKLAPKAEAKTVQOE- 281  
DB 435 DKDTPAVFTTISLSLKHKAELAKLAHTTVSRNVVAEANA---IEKEVVPEKAEKKVKEED 491  
QY 282 ---VAESVSF---GELPSHMDTKHIERIPWASEQAOQTVSQHLHAGNLSELGNLNNM--N 333  
DB 492 VKAAVEEKKEKPKGKLP--WKIEKLEK-PVDTK-----SENHELKVKVLD 534  
QY 334 RDLAPHLRLREVSDFRSEPHSPISFLLEK---AIRMGYLSLPELLREMSQ---NGDA 387  
DB 535 KERALLVESEIKNTAEETKPKVE-SFKSEETVAID---DMPALEEESAEEKKETTGE 589  
QY 388 LSTFNAAGLNHLDOVLLPEVSTFTVGLIESPQTQAKPSVSDPRSVZEHVSTQSPVDQTQ- 446  
DB 590 TTT-----EAAVETTEASETPKPEAKPELLS--NLEDVLTLTTP-ETET 630  
QY 447 ---SKQDOKPOSSATSA 460  
DB 631 IEGSEREPTTSAPAA 647

## RESULT 6

G83634  
Hypothetical protein PA0082 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: G83634  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; B.  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Foiger, K.R.; Kas, A.; Larbig, K.; Lim  
.: Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path  
A:Reference number: AB2950; MUID:20437337; PMID:10984043  
A:Accession: G83634  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-344 <STO>  
A:Cross-references: GB:AE004447; GB:AE004091; NID:G9945902; PIDN:AAG03472.1; GSPDB:GN00  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA0082

Query Match 6.0%; Score 141.5; DB 2; Length 344;

Best Local Similarity 21.1%; Pred. No. 0.069;  
Matches 99; Conservative 70; Mismatches 144; Indels 119; Gaps 18;

QY 15 LSDDSICGVYKLEKSAFRPLRNFVNAQTALRKLQ-NPSADDERALQEAACLNKWKILS 73  
DB 11 VSPDSPCGDDL-----EYDAAFLERLTAQGPQRMQMGDAVLPAPPEWPRVR 58  
QY 74 DSIYEQFSKTRTDIELISWFAAQFLDITLESANSL-ENLADLSEKHLNPLVPET 133  
DB 59 ALASELFGK-SKDLRVANLLQSNVALD-GLDGLADGLLVRELLGQYWDGVYPL----- 111  
QY 134 LKSDDDKGEREQADAKYKAPQLVGDSESSILYAPVLQ-----PLVGEVTF--FDQS 187  
DB 112 LQADDD-----NDPTFRINALTGLVAE-----PLLQVWALFVRSRAFGVPNLR 157  
QY 188 AERKGEISOLKSLMTTVAERFAIOFKMENAKRCVTQDLRSALYSTKCHSLGQSQTNF 247  
DB 158 ALNAAQLQFPAS---ETLSPEQIAGAFADADA-----DALAATRALDGAQEHAL 204  
QY 248 GFAKSLLTRVENA-----LVHL-----SGIKLAPKAEAKTVQOEVAES 285  
DB 205 AIESGVAERVGSAGQGLDGLRQLRQALQVFDLYPQAGESLAPGAEA-VADEQVGAA 263  
QY 286 SVSEGEPLPSHMDTKHIERIPWASEQAOQTVSQHLHAGNLSELGNLNNMNRDLAFHLREVS 345

Db 264 PVAATAA-----PRAS-----GEIA-----NREDVLRQLDRLL 293  
 QY 346 DYFROSEPHSPISFLEKAIKRWGLSLPELLREMSQNGDALSTIFNAAGLNHLDOVLL 405  
 Db 294 EYVVRHESPVPVLLKRAKTLVADFAIVRNLPD-----GISQFETLRG 340  
 QY 406 PE 407  
 Db 341 PE 342

RESULT 7  
 T16927  
 hypothetical protein T23F2.2 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
 C:Accession: T16927  
 R:Du, Z.  
 submitted to the EMBL Data Library, October 1995  
 A:Description: The sequence of C. elegans cosmid T23F2.  
 A:Reference number: Z18608  
 A:Accession: T16927  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1510 <DUZ>  
 A:Cross-references: EMBL:U39649; NID:gl049370; PID:gl049372; PIDN:AAA80384.1; CESP:T23F2  
 C:Genetics:  
 A:Gene: CESP:T23F2.2  
 A:Introns: 20/2; 81/1; 119/2; 211/3; 236/3; 290/2; 344/1; 396/3; 439/3; 465/3; 556/3; 60

Query Match 6.0%; Score 141.5; DB 2; Length 1510;  
 Best Local Similarity 20.0%; Pred. No. 0.56;  
 Matches 112; Conservative 89; Mismatches 168; Indels 191; Gaps 26;

QY 9 BQLKPLSDSDSICGVYKLEKSAFRLNENFVAQATLRLKLSQNPSPADDERDALQEA 68  
 Db 619 QQLSEHSD-----FLSTQOT-LHDLKKEYSAKNTTL-----VDKFKVEEILLAK 663  
 QY 69 WKILSDSLYEQFSK---TTDRDIELISWFAAQFLDITLESAAANSLEWLADESEKHW 124  
 Db 664 TELV-DALTKLENIRKDOTRELS-----LKQSERDQYKSLSEMTFAEK----- 708  
 QY 125 LNPVLVETLKSDDKQKEREQADAKV---AFFQ-----LVGDS----- 161  
 Db 709 -VPIEAELIQLSKDK---NEITARKHQDQVPEDELAKLNDMSNMKIKKRDYLTET 763  
 QY 162 --ESSILYAPVLQPLVGVETFFDFOAER-KGEISQKSMITTTVAQERFAIQFMEN 218  
 Db 764 RANESMI-----BRKLKLEISGLKLDLENQKQMAHLQ---KSEL 798  
 QY 219 AKRCVTQDLRLSALVS-----TKCHSLGSGSTNFGPAK----- 251  
 Db 799 EKILLSIDHVSQLSRVNSQRDVECAIPROINKVVGCKPVNKNKETTIEKGAFLDEN 858  
 QY 252 -----SLLTRVENALVHL-----SOIKLAPKAEAKTVQEVVAES 286  
 Db 859 EERLIRICAELETTTRQVTVLQKLSIIQQQSQKIKKRIAVVEDSNKNTVHTEDLES 918  
 QY 919 MKEVEL---XNTELMERID-SLEAERFVASSIEKSRIOKLNVNFDNLKQKLDNDMSNYSK 974  
 QY 335 DLAF-----HLREVSDFRQSEPHSPISFLEKAIKRWGLSLPE-LIREMMSQNG- 386  
 Db 975 EKQWLQWRISNLEKONSELOKIQPSS-----EKSLE-----SLNKGTLRKMTSEPDFGD 1024  
 QY 387 ALSTIFNAAGLNHLDOVLLPEVSTPTVGIESPQ-----TPQAKPSVSDPR 431  
 Db 1025 DMSTEGDGASTNSADPWSVETAPVLKSLKSPQFSQLADVLNLRVSDLEQVLTIEEPE 1084  
 QY 432 SVEEHVSQTSFVDTQSKQDQ 451

Db 1085 AAKQEEPQMSLEKSTNKNVQ 1104

RESULT 8  
 E98012  
 hypothetical protein smc [imported] - Streptococcus pneumoniae (strain R6)  
 C:Species: Streptococcus pneumoniae  
 C>Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001  
 C:Accession: E98012  
 R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E  
 e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M  
 y, P.; Sur, P.M.; Winkler, M.E.  
 J. Bacteriol. 183, 5709-5717, 2001  
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;  
 A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
 A:Reference number: A97872; MUID:21429245; PMID:11544234  
 A:Accession: E98012  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1179 <KUR>  
 A:Cross-references: GB:AE007317; PIDN:AAK99929.1; PID:gl5458753; GSPDB:GN00174  
 C:Genetics:  
 A:Gene: smc

Query Match 5.7%; Score 134.5; DB 2; Length 1179;  
 Best Local Similarity 21.6%; Pred. No. 1.1; Mismatches 80; Indels 105; Gaps 24;  
 Matches 100; Conservative 80; Mismatches 80; Indels 105; Gaps 24;

QY 5 KQIQLKPLSDSDSICGVYKLEKSAFRLNENFVAQATLRLKLSQNPSPADDERDALQEA 64  
 Db 674 KPELQLQKEIAADS---ASLGSEBAALKTLQDQMAALTERLEAIK---SOGQQAIRIQEQ 727  
 QY 65 CLNKKWILSDSLYEQFSKTTDRDIELISWFAAQFLDITLESAAANSLEWLADESEKHW 124  
 Db 728 GLS-----LAYQOTSOQVEETL-WKQEE-EIDRLSEG-----DWQAD-KEKQES 772  
 QY 125 L-----NPVLVETLKSDDKQKEREQADAKVKAFFQVGVGSESSILYAPVLQPL 176  
 Db 773 LATIASQKLEAEIEIKSNKALQERYQ-----NLQEEVAQARLLATEL 818  
 QY 177 VGEVTFPFOAERKGEISQKSMITTTVAQERF-AIQFMENAKRCVTQDLRLSALVS 234  
 Db 819 QGQKE-YEVADIERLG-----KELDNLEIEQEIQRLQEKVDNLEKVDTE-----LLS 866  
 QY 235 TKCHSLGSGSTNFGPAKSL-LTRVENALVHLSG-----IKLAPKAEAKTV 278  
 Db 867 QCAESKTQKTNLQGLIRKQFELDDIEGLDDIASHLDQARQONBWRKQFAEAK-- 924  
 QY 279 BOEVAESSVSGELPSPMDTKHIERIPMAEQATVQSHLHAGNLS--ELGNLNNMRDL 336  
 Db 925 KEKVSERL-----RHLQNLTDQYQTSYTEALEKAHELENLNLAPQEVQDLKAIKRS 977  
 QY 337 AFHLLREVSDFRQSEPHSPISFLEKAIKRWGLSLPELLREMSQNGA---LSTIEN 393  
 Db 978 GPNVLEALDQY---EEVNRDLDFL--NSQRDILSAKLLLETITENDEVKERFKSTFE 1032  
 QY 394 AA-----GLNHLDOVLLPEVSTPTVGIESPQTPOAK 424  
 Db 1033 AIRESEFKVTFKQMGFGGQAD-LILTEGDLTLTAGVEISVQPPGK 1074

RESULT 9  
 T18302  
 apSB protein - Emericella nidulans  
 C:Species: Emericella nidulans, Aspergillus nidulans  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 11-May-2000  
 C:Accession: T18302  
 R:Suellmann, R.; Sievers, N.; Galetzka, D.; Robertson, L.; Timberlake, W.E.; Fischer, R.  
 Mol. Microbiol. 30, 831-842, 1998  
 A:Title: Increased nuclear traffic chaos in hyphae of Aspergillus nidulans: molecular ch  
 A:Reference number: Z18870; MUID:99140441; PMID:10094631  
 A:Accession: T18302  
 A:Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA  
A;Residues: 1-1051 <SUE>  
A;Cross-references: EMBL:AJ003163; NID:e1286086; PID:e1286087; PIDN:CAA05918.1  
C;Genetics:  
A;Gene: apsB  
A;Introns: 160/3

Query Match 5.7%; Score 134; DB 2; Length 1051;  
Best Local Similarity 19.2%; Pred. No. 1;  
Matches 86; Conservative 62; Mismatches 134; Indels 166; Gaps 17;  
QY 7 QIQLKPLSDSDSICGVYKLEKSAFRLNENVAQTALRKLSONPSADERDALQ---- 62  
DB 416 KIECOR-LTED-----LFTQDNLRALQAEKRSASEGIIRLEE-----DAQNNLQRYKA 464  
QY 63 -----EACLNKWLSDSLVEQFSKTRDIELISWFAAQFLDITLLESAANSLEWLADL 117  
DB 465 VQOELEOCNEEMESLEKSLYEANSKVQR-----LTVQIESQNEIAFLRE- 509  
QY 118 SEKHWDHLNPLVETLKSDDKGRERQADAKVAFQVLGDSSESSILYAPVLQPLV 177  
DB 510 -----EQDGDKIK-----IGDSELSKTYRMSLQ----- 533  
QY 178 GEVTFDFQSAERKGBISQLKSLMTTVAQERFAIQF-----KMNKRCVDTQDLRLSALV 233  
DB 534 -----SEKDKYKELEGL-----AEERYQREVVGSKKEQEVQRMWELNREVSA 578  
QY 234 STRCH-----SLGQSTNFGFAKSLTRVENALVELSQ-----IKLAPKAERK 276  
DB 579 KEECKLKNLKAQAEITNTWKRUTDLENNLRATLGLDTSRSSLIANIMKLQKELEST 638  
QY 277 TVQOEVAESSVSEGE-----LPSH-MDKHIERIPMASEQA-----QTVSQH 317  
DB 639 ALELESTRTDEKETLLNRDALLSHGLSRLKSELLERERQARADKQSFQALKSH 698  
QY 318 LHA-----GNLSELGNLNNMRDLAFHLLREVSDYFRQSEPHSPISFLLEKAIKRWGY 369  
DB 699 HQASRTITQNNSRILEENARNQDRK-----R- 726  
QY 370 LSLPELLREMMSEONGDALSTIFNAAGL 397  
DB 727 TSLEQQFQEQLEQINERNMSLLTIWKRLSGM 754

## RESULT 10

T22976  
hypothetical protein F59A2.6 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C;Accession: T22976; T23157  
R;Lightning, J  
submitted to the EMBL Data Library, June 1994  
A;Reference number: Z19645  
A;Accession: T22976  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1133 <W1L>  
A;Cross-references: EMBL:Z34801; PIDN:CAA84332.1; GSPDB:GN00021; CBSP:F59A2.6  
A;Experimental source: clone F59A2  
R;Burton, J.  
submitted to the EMBL Data Library, October 1995  
A;Reference number: Z19700  
A;Accession: T23157  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1133 <W12>  
A;Cross-references: EMBL:Z66514; PIDN:CAA91344.1; GSPDB:GN00021; CBSP:F59A2.6  
A;Experimental source: clone K01A11  
C;Genetics:  
A;Gene: CBSP:F59A2.6  
A;Map position: 3  
A;Introns: 13/2; 43/3; 107/3; 413/3; 492/3; 567/3; 635/3; 710/3; 738/3; 795/2; 1008/3; 1

Query Match 5.5%; Score 129.5; DB 2; Length 1133;  
Best Local Similarity 21.3%; Pred. No. 2.2;  
Matches 110; Conservative 77; Mismatches 184; Indels 145; Gaps 23;  
QY 9 EQLSKPLSDSDSICGVYKLEKSAFRLNENVAQTALRKLSONPSADE-----RDALQEA 64  
DB 116 EQLAK-----AMEKLNSEQNIIDVETKLEQ--SEBEVLAARGAIOE- 155  
QY 65 CLNKWKILSDSLVEQFSKTT---RDIELISWFAAQFLDITLLESAANSLEWLADLSKH 121  
DB 156 -----LTKLEESEKETSTAKTELEAVS-----KKLDSSETSLKEFSDMIE- 196  
QY 122 WDHLNPLVETLKSDDKGRERQADAKVAFQVLGDSSESSILYAPVLQPLVGSVT 181  
DB 197 -----AMKIQLINECKQKQDAVELLKQKLEBEKNMSDVE-----VKQLLEST 241  
QY 182 PFDF-----QSAERKGEISQLK-----SMLTTTVAQERFAIQ---FKMENA 219  
DB 242 TSEMKAHAAEIVKQLEBAQSSILENKDAENERNLKTALDESDESSAIEITKQMEAA 301  
QY 220 KRCV-----TQDLRLSALVSTKCHSLGSO-----SINFGFAKSLITVENAL 261  
DB 302 KKELEASEKSEKSELEQMDRLQ-----KVHAGQEDIOKLOKTELEMAKIAKSTEDKL 356  
QY 262 V-HLSGKILAPKAAKTVEQEV-----ABSSVSEGLPSHMDTKHIERIPMASEQAOT 313  
DB 357 AREQLAGLENAKBDLVVEBEKHTGIQRAQGLDDAEKVKVLKEQLERAQSALESSOE 416  
QY 314 VSQHLHAGNLSEL-GNLNNMR-----DLAPHLLEVSDFRQSEPHSPISFLLEKAIK 366  
DB 417 LASSQKADKIOELEKELQNAQKRSESELETANEMVRSITATLENS--NSETILKQK--- 471  
QY 367 NGYLSPELLREMMSEONGD-ALSTIFNAAGNLHDQVLLPEVSTPTVGIESPQTPQAKP 425  
DB 472 -----LETLDRELQARQOTEKALTEINVLTS-----LAEKQOTAIQNLQITQIM 520  
QY 426 VSYSDPRSVEEHVQSPTVDTSKQDKQPSQATSAL 461  
DB 521 EYEKEEKEVE-----LVKVLQQAQSSSSAAEAL 549

## RESULT 11

T42977  
large tegument protein - ateline herpesvirus 3 (strain 73)  
C;Species: ateline herpesvirus 3  
A;Variety: strain 73  
C;Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jan-2000  
C;Accession: T42977  
R;Albrecht, J.C.; Fleckenstein, B.  
submitted to the EMBL Data Library, August 1998  
A;Description: Primary structure of the herpesvirus ateles genome.  
A;Reference number: Z22274  
A;Accession: T42977  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-2471 <ALB>  
A;Cross-references: EMBL:AF083424; PIDN:AAC95568.1  
A;Experimental source: strain 73

Query Match 5.4%; Score 127.5; DB 2; Length 2471;  
Best Local Similarity 19.1%; Pred. No. 9.2;  
Matches 121; Conservative 90; Mismatches 187; Indels 237; Gaps 31;  
QY 1 MPLSKHQIEQLSKPLSDSDSICGVYKLEKSAFRLNENVAQTALRKLSON-----PS 54  
DB 660 LPTKVLIEIQPLEK-----SKVFLEKLPVEQKLTTEVNEILLTLLHNKIKODATEILPV 712  
QY 55 ADERDALQ-----EACLNKWLSDSLVEQFSKTRDIELISWFAAQFLDITLLE 105  
DB 713 PDFTILKNIQSTLQLLHCTVDTLMI--DK--EVIGSTIQQLSYIGWEV----- 757  
QY 106 SAANSLEWLADLSKHXWDR--LNPVLPVETL---KSSDDKGRERQADAKVAF- ---Q 156



Db 132 SKPSLIRDAIKHRAKIDSVTEAKK-----KLDLYKEKNAHISSEDIHTGDFDSFM 185  
QY 255 TRVENALVHLSGI---KLAKAEAKTVQEVAESSVSE-----GEL 292  
Db 186 NKQGGAKPLEATSEALSAAES-NIINKLAKSVPTTFIDFKDDPMKLAKETEBFGKI 244  
QY 293 PSHMDTKH-----IERIPMASEQ-----AQTVSOHLHAG----- 321  
Db 245 SINEYSKSKQFLLEHLPIISEQKDALMKWKAFFYQLHGDDKMTLQVHQSELMAIYEIY 304  
QY 322 -----NLSLGNLNMNEDLAFHLLREVSDYFRQSEPHSPISFLEKAIWGYLSLP 373  
Db 305 DMKIPYLNPMELSVNIM-----FFEKV--IFNKDKPMGKESFL--RSVQKFLHIQ 353  
QY 374 E-----LLREMMSEONGDALSTIFNAAGLNHLDQVLLPEVSTPTVGIESPQ 419  
Db 354 KRSKILQOEEMDESNAEGVETI---OLKSLDDSTELVNLDPFNSKDPE 399  
RESULT 14  
T22235  
hypothetical protein F45G2.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T22235  
R:Lindsay, S.  
submitted to the EMBL Data Library, March 1997  
A:Reference number: Z19535  
A:Accession: T22235  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1287 <N1>  
A:Cross-references: EMBL:Z99392; PIDN:CA807612.1; GSPDB:GN00021; CESP:F45G2.3  
A:Experimental source: clone F45G2  
C:Genetics:  
A:Gene: CESP:F45G2.3  
A:Map position: 3  
A:Introns: 59/2; 117/3; 153/3; 180/2; 233/2; 341/3; 393/1; 424/2; 551/3; 664/3; 734/3; 8

Query Match 5.4%; Score 125.5; DB 2; Length 1287;  
Best Local Similarity 21.4%; Pred. No. 4.9;  
Matches 90; Conservative 79; Mismatches 167; Indels 85; Gaps 18;  
QY 7 QIEQLSKPLSDSDSICGVYLKLSAPRPLNEFNVAQTALRKLSONPSADERDALQEA 66  
Db 259 EIEELNRLKAESESNEYENAKKASELERLR-----EKKEKDEKERRMCMAEANEA-- 310  
QY 67 NKWKILSDSYEQPSKTRTDIELISWFVAAQFLDITLLESAANSLEWLADLSEKHWDHLN 126  
Db 311 NQHKTOEKLKQDQISTLTQTFEKVSAQRKAQEQMNAELVDEVASFKQKAERAEQKKL- 369  
QY 127 PVLVETLKSDDDK-GKERQADAKYKAPFQIVGDSEESSILYAPVLPVLP-VGEVTFPD 184  
Db 370 -----VEDLSDDDKLAEEKKANNEHVHKNKLEQG-----LKATQTLTALUKHEFD 419  
QY 185 FQSAERKEISQLKSMLTITTTVAQERFAIQFQKME-NAKRCVTDRLSALVSTKCHSLGSQ 243  
Db 420 VDCKRSEIGEELK-----LKAQGDANLISKQLAMLRKCSIRIEELEEDLEE----- 467  
QY 244 STNFGFAKSLTRVENALVHLSGIKLAAPKAEAKTVQEVAESSVSGEIPSHMDTKHIER 303  
Db 468 -----RKLRMKAERQFNL-----RSEYEVLOEQMAEAS---GQLTAE---AHINK 507  
QY 304 IPWASEQAQVSOHLHAGNLSE-----LGNLNNMNR---DLAPHLLEVS 346  
Db 508 V--RAEVSNLRLDQKRLNHNHAYISDLNMQVATVNNRLNLSQQFYSFNFNCFFFL 565  
QY 347 YFRQSEPHSPIS-FLLEKAIWGYLSLPELLEMMSEONGDALST-----IFNAAGL 397  
Db 566 FFGPSIYDKDPIHLFOIRKIIK--SPVTVIFHELISRESVPTVASLCIQAQTIIFNGASL 622  
QY 398 N 398

Db 623 N 623

## RESULT 15

A47297

myosin heavy chain form B, nonmuscle - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C:Date: 22-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 02-Feb-2001

C:Accession: A47297; A55441

R:Bhatia-Dey, N.; Adelstein, R.S.; David, I.B.

Proc. Natl. Acad. Sci. U.S.A. 90; 2856-2859, 1993

A:Title: Cloning of the cDNA encoding a myosin heavy chain B isoform of Xenopus nonmuscle

A:Reference number: A47297; MUID:93219383; PMID:8464900

A:Accession: A47297

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-1992 &lt;BHA&gt;

A:Cross-references: GB:I09740; NID:g214623; PIDN:AAA49915.1; PID:g214624

A:Experimental source: XTC cells

A:Note: sequence extracted from NCBI backbone (NCBIP:128722)

J. Kelley, C.A.; Oberman, F.; Yisraeli, J.K.; Adelstein, R.S.

J. Biol. Chem. 270, 1395-1401, 1995

A:Title: A Xenopus nonmuscle myosin heavy chain isoform is phosphorylated by cyclin-p34

A:Reference number: A55441; MUID:95138137; PMID:7836406

A:Accession: A55441

A:Status: preliminary

A:Molecule type: protein

A:Residues: 198-232 &lt;REL&gt;

C:Superfamily: myosin heavy chain; myosin motor domain homology

C:Keywords: nucleotide binding; P-loop

F:88-787/Domain: myosin motor domain homology &lt;MMOT&gt;

F:178-185/Region: nucleotide-binding motif A (P-loop)

Query Match 5.4%; Score 125.5; DB 2; Length 1992;  
Best Local Similarity 21.2%; Pred. No. 9.1;  
Matches 107; Conservative 78; Mismatches 189; Indels 131; Gaps 20;

QY 1 MPLSKHQIQEQLSKPLSDSDSICGVYLKLSAPRPLNEFNVAQTALRKLSONPSADERDA 60  
Db 929 LAIKQEMEEILRDLE-----IRVEEBEERNQVLQNEKKVQTHVQDLEE--QDDEEA 980  
QY 61 LQEAELNK-----WKILSDSLY--EQFSKTRTDIELISWFVAAQFLDITLLESAANS 111  
Db 981 AQKQLKLVETAKIKKVEEDILVLEQNSKFLKEKLEERIAESTQLAEEBEKAKNL 1040  
QY 112 EMIALSKHNDHLPVLPVETLKSDDDKGEREQADAKVKAFFOLVGDSEESSILYAPV 171  
Db 1041 AKLKNQEMMISDLF-----ERLKEEKTRQELKAKRK----- 1074  
QY 172 LQLPLVGEVTFDFQSAERKGEISQLKSMLTITTTVAQERFAIQFQKMEAKRCVTO----- 225  
Db 1075 -----LDGTTTDFQDQIAELQAIQELKQL-----AKKEEIQALARGDEVLQNNILK 1126  
QY 226 -LDRLSALVSTKCHSLGSQSTNFGFA-----KSLTRVENALVHLSGIKLAPKA 273  
Db 1127 LVRELOAQIAELQEDLESEKASRNKAERKQKRDLSSELEALKTELEDTL-----DTTAAQ 1181  
QY 274 EAKTV-EQVAESSVSGEELPSH-----MDTKHIERIPWASEQAQTVSOHLHAGN 322  
Db 1182 ELRTKREQVAVELRKSIIEETRNHEAQEMRQROATALEELSEOLEQAKKPKVNLKNN 1241  
QY 323 LSELGNLNMNRDLAFHL-----LFEVSDYFRQ-----SEPHSPISFLEKAIWGYLS 371  
Db 1242 QS-----LESNDKELATEVKSILQMKAESEYKRRKLEGGVQVQLHAKV---LE----- 1285  
QY 372 LPPELLREMMSEONGDALSTIFNAAGLNHLDQVLLPEVSTPTVGIESPOTPOAKSVSDPR 431  
Db 1286 -GDLRLADMKVEKSKLQNELENVSS-----LLEAEKKGKIL-----AKDVA 1326  
QY 432 SVEEHVSQTSVPVDTQSKODQKPPQS 456  
Db 1327 SMESQLQTOELLQBEETQKLNQSS 1351

Tue Jul 6 09:50:14 2004

us-10-780-347-2.rpr

Page 8

Search completed: July 3, 2004, 02:46:49  
Job time : 30 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2004, 22:59:50 ; Search time 19 Seconds  
(without alignments)  
1269.867 Million cell updates/sec

Title: US-10-780-347-2

Perfect score: 2343

Sequence: 1 MPLSKHQIEQLSKPLGDSI.....DTQSKQDQKQSSATSALSW 463

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	152	6.5	2245	1 MYSJ_DICDI	P54697 dictyosteli
2	134	5.7	1051	1 APSB_EMENI	O60039 emericella
3	129.5	5.5	790	1 BEAL_MOUSE	O8b166 mus musculu
4	128.5	5.5	877	1 CAPP_VIBVU	Q8dcm2 vibrio vuln
5	126	5.4	506	1 CC37_YEAST	P06101 saccharomyc
6	125	5.3	793	1 REGA_DICDI	Q23917 dictyosteli
7	125	5.3	1194	1 SMC3_SCHPO	O42649 schizosacch
8	124.5	5.3	1940	1 MVH3_HUMAN	P11055 homo sapien
9	124.5	5.3	1940	1 MVH3_RAT	P12847 rattus norv
10	124	5.3	866	1 MYSJ_SCHMA	P06138 schistosoma
11	123.5	5.3	867	1 DPOL_RICFR	O05949 rickettsia
12	123	5.2	1935	1 MYH7_HUMAN	P12883 homo sapien
13	123	5.2	1935	1 MYH7_RAT	P02564 rattus norv
14	123	5.2	2022	1 ANTI_ONCVO	P21249 onchocerca
15	122.5	5.2	1220	1 LBN_MOUSE	O8k192 mus musculu
16	122.5	5.2	2469	1 TEGU_HSVSA	O01056 herpesvirus
17	122	5.2	1934	1 MYH7_MESAU	P13540 mesocricetu
18	121.5	5.2	1679	1 GCC2_MOUSE	Q6ch93 mus musculu
19	121.5	5.2	3259	1 GOB1_HUMAN	Q14789 homo sapien
20	121	5.2	1357	1 KTN1_HUMAN	Q86up2 homo sapien
21	120.5	5.1	999	1 OXRP_HUMAN	Q9y411 homo sapien
22	120.5	5.1	1938	1 MVH4_RABIT	Q28641 oryctolagus
23	120	5.1	667	1 M110_CAEEL	P24400 caenorhabdi
24	119.5	5.1	1066	1 KL61_DROME	P46863 drosophila
25	119	5.1	1935	1 MYH3_PIG	P79293 sus scrofa
26	119	5.1	2104	1 MYS3_SCHPO	O14157 schizosacch
27	118.5	5.1	1920	1 PCT2_MOUSE	P48725 mus musculu
28	118.5	5.1	1937	1 MYH8_HUMAN	P13535 homo sapien
29	118	5.0	1979	1 TRIA_HUMAN	O15643 homo sapien
30	117.5	5.0	539	1 YAH2_YEAST	P39705 saccharomyc
31	117.5	5.0	1427	1 REST_HUMAN	P30622 dictyosteli
32	117.5	5.0	2116	1 MYS2_DICDI	P08799 dictyosteli
33	117.5	5.0	2611	1 BP1E_MOUSE	Q912u8 mus musculu

34	117	5.0	1679	1 YIO9_YEAST	P40457 saccharomyc
35	117	5.0	2779	1 LVA_DROME	Q8nes1 drosophila
36	116	5.0	1583	1 GCC2_HUMAN	Q8awj2 homo sapien
37	116	5.0	1616	1 P200_MYCGE	Q49429 mycoplasma
38	116	5.0	1966	1 MYSB_CAEEL	P02566 caenorhabdi
39	115.5	4.9	1690	1 C190_DROME	Q9vje5 drosophila
40	115.5	4.9	2468	1 MAPB_HUMAN	P46821 homo sapien
41	115.5	4.9	3660	1 DMD_CHICK	P11533 gallus gall
42	115	4.9	1312	1 RAS0_YEAST	P12753 saccharomyc
43	114.5	4.9	955	1 KINL_LEICH	P46865 leishmania
44	114.5	4.9	2230	1 GOR4_HUMAN	Q13439 homo sapien
45	114.5	4.9	5327	1 MACF_MOUSE	Q9qxx0 mus musculu

#### ALIGNMENTS

RESULT 1

MYSJ\_DICDI

ID MYSJ\_DICDI STANDARD; PRT; 2245 AA.

AC PS4697;

DT 01-OCT-1996 (rel. 34, Created)

DT 01-OCT-1996 (rel. 34, Last sequence update)

DT 15-MAR-2004 (rel. 43, Last annotation update)

DE Myosin IJ heavy chain.

GN MYOJ.

OS Dictyostelium discoideum (Slime mold).

OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.

OX NCBI\_TaxID=44689;

RN (1)\_

RP SEQUENCE FROM N.A.

RC STRAIN=AX3;

RX MEDLINE=96215148; PubMed=8636147;

RA Hammer J.A. III, Jung G.;

RT "The sequence of the dictyostelium myo J heavy chain gene predicts a

RT novel, dimeric, unconventional myosin with a heavy chain molecular

RT mass of 258 kDa.";

RL J. Biol. Chem. 271:7120-7127(1996).

RN (2)\_

RP SEQUENCE OF 1-1021 FROM N.A.

RX MEDLINE=97039016; PubMed=8884597;

RA Peterson M.D., Urioste A.S., Titus M.A.;

RT "Dictyostelium discoideum myoJ: a member of a broadly defined myosin

RT V class or a class XI unconventional myosin?";

RN J. Muscle Res. Cell Motil. 17:411-424(1996).

RN (3)\_

RP SEQUENCE OF 182-298 FROM N.A.

RX MEDLINE=95023928; PubMed=7937787;

RA Titus M.A., Kuspa A., Loomis W.F.;

RT "Discovery of myosin genes by physical mapping in Dictyostelium.";

Proc. Natl. Acad. Sci. U.S.A. 91:9446-9450(1994).

CC -!- FUNCTION: Myosin is a protein that binds to actin and has ATPase

CC activity that is activated by actin.

CC -!- SUBUNIT: Homodimer.

CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.

CC -!- SIMILARITY: Contains 1 dilute domain.

CC -!- SIMILARITY: Contains 3 IQ domains.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; U42409; AAA85186.1; -

CC EMBL; L35322; AAA79858.1; -

CC PIR; T18278; T18278.

CC KSSP; P08799; 1MND.

CC DictyBase; DDB0185050; myoJ.

CC InterPro; IPR002710; DIL.

CC InterPro; IPR000048; IQ\_region.

CC -----



DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Early endosome antigen 1 (Fragment).  
 GN ZEAL.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Brain;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Boto H., Kondo S.,  
 RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojohori T.,  
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schirral L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Brad D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
 RA Dalla E., Dragani T.A., Fletcher C.P., Forrest A., Frazer K.S.,  
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
 RA Gimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Mgiotti D.R., Maltais J., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Perte G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Sadelin A., Schneider C., Sample C.A., Setou M., Shimada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Varardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
 RA Yuan Z., Zavalan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 RA Hiroxane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.,  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs";  
 RL Nature 420:563-573(2002).  
 CC -!- FUNCTION: Binds phospholipid vesicles containing  
 CC phosphatidylinositol 3-phosphate and participates in endosomal  
 CC trafficking (By similarity).  
 CC -!- SUBUNIT: Homodimer. Binds STX6. Binds RAB5A, RAB5B, RAB5C and  
 CC RAB22A that have been activated by GTP-binding (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic and peripheral membrane protein  
 CC associated with early endosomes (By similarity).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AK046231; BAC32647.1; .  
 DR MGD; MGI:2442192; Beal.  
 DR KW Coiled coil.  
 FT DOMAIN 1 789 COILED COIL (POTENTIAL).  
 FT DOVAIN 180 550 GLN/GLU/LYS-RICH.  
 FT NON TER 790 790  
 SQ SEQUENCE 790 AA; 90539 MW; A167AE02CC3CABTF CRC64;  
 Query Match 5.5%; Score 129.5; DB 1; Length 790;  
 Best Local Similarity 21.0%; Pred. No. 1.2;  
 Matches 104; Conservative 85; Mismatches 175; Indels 131; Gaps 21;  
 5 KHQTEQLSKLSDSDSICGVILKESAFRLRNEFNVAQTALRKL-----QNPSSA--- 55  
 324 QHLEKSKQHQBQ-----ALQSSATKLRAQNDELQVLKQIGDKDKQKIQNLEALQ 377

QY  
 DB

QY 56 -----DERDAL-----QEACLNKKWILSDSLVEQFSKTRTDIELISFWVAA 96  
 DB 378 KGESVSLLKEREEDLYAKIQAGEGETAVLNQLOEKHNLQCOLTLQTEKLNQS----- 432  
 QY 97 QFLDTTLESAAANSLWLAJSEKWHDL-----NPVLVET-----LKSDDDKGRER-ROA 147  
 DB 433 -----ESHKQABENLHDVQEQKAHLRAAQDRVLSLETSSVLSQLNESKEKVSQ 484  
 QY 148 DAKVKAFTQVGDSESSILYAPVLQPLVGEVTFDFQSAERKGEISQLKSMLTVAQ 207  
 DB 485 DQIKAKTELLLSAE-----AKAAQRAADLQNLDTA--- 516  
 QY 208 ERFATQFVXENAKRCVTQLRLSLVSTK---CHSLGSCSTNFGPAK-SLITRVENALVH 263  
 DB 517 -QHAIQDQKQELNKVSVQLDQLTAKFQEKHECHIQLESHLKDHEKHLISLEQKVEDLSEH 575  
 QY 264 LSGIKLAPKAEAKTVEQEAESSVSEGLPSHMDTKHTERIPMASEQATVUS-QHLHAGN 322  
 DB 576 IKKLE-ADALEVKASKQALOSLOQORQLSTDLRLNAELRELOEQEVEVSCTKLDLQN 634  
 QY 323 LSELGNLNMNRDLAFHLLREVDVDFRQSEPHSPISFLLEKAIKRWGY-SLPELLREMMSE 382  
 DB 635 KSEI--LENIKQTLT-----KKEEENVLKQEFELKSDSKTKQHKEL----- 674  
 QY 383 QNGDAL-STIFNMAAGNLHLDQVLLDEVSTPTVGIESPOTPAKPSVSPRSVEHVQS 441  
 DB 675 ---GDRMQAAVTETLTAQKQDALLAELST-----TKELSKVSD--SLKNSKSEF- 720  
 QY 442 PVDTSKODQKPOSS 456  
 DB 721 -----EKENQKXAA 730  
 RESULT 4  
 CAPP\_VIBVU STANDARD; PRT; 877 AA.  
 ID\_O6D62;  
 AC 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Phosphoenolpyruvate carboxylase (EC 4.1.1.31) (PEPCase) (PEPC).  
 GN PPC OR VW11369.  
 OS Vibrio vulnificus.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=672;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CMCP6;  
 RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,  
 RA Choy H.E.;  
 RT "Complete genome sequence of Vibrio vulnificus CMCP6";  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Forms oxaloacetate, a four-carbon dicarboxylic acid  
 CC source for the tricarboxylic acid cycle.  
 CC -!- CATALYTIC ACTIVITY: Phosphate + oxaloacetate = H(2)O +  
 CC phosphoenolpyruvate + CO(2).  
 CC -!- PATHWAY: Tricarboxylic acid cycle.  
 CC -!- SIMILARITY: Belongs to the PEPCase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AB016801; AAC09818.1; .  
 DR HAMAP; MF 00595; .; 1.  
 DR InterPro; IPR001449; PEPCase.  
 DR Pfam; PF00311; PEPCase; 1.

```

DR PRINTS: PR00150; PEPCARBOXYLASE.
DR PROSITE: PS00781; PEPCASE.1; 1.
DR PROSITE: PS00393; PEPCASE.2; 1.
KW Lyase; Carbon dioxide fixation; Tricarboxylic acid cycle;
KW Complete proteome.
FT ACT SITE 138 BY SIMILARITY.
FT ACT SITE 544 BY SIMILARITY.
SQ SEQUENCE 877 AA; 98900 MW; 14CFD7BB65CA779C CRC64;

Query Match 5.4%; Score 128.5; DB 1; Length 877;
Best Local Similarity 23.2%; Pred. No. 1.6; Mismatches 129; Indels 151; Gaps 28;
Matches 105; Conservative 67;

QY 3 LSKHQIEQLSPLDSDGICGYLKLKSAFRLNEFNVAQTALRK-----LSQPSAD 56
DB 91 ISRHDAHVCEP---DAINTLFAKLQNGI---NKLDTAQ-AIRELNIELVLTAHPT 142
QY 57 ERDAL-----OEACLNKWLSDSLYEQSKTTRDIELI---SW---FVAAQFLDITL 104
DB 143 TRRTMINKLVKINECLSKLE-LSLSYKBRHKTEKLEQLIAQSHSDVIRKQ--RPTPL 199
QY 105 ESA-----ANSL-----EWIADLSEKHWDHNPVLPVETLKSDDDKGKEREQADAKVK 152
DB 200 DEAKGFAVENSWEAVDFELREDEKLDYLDQGLPI-----DAR-- 241
QY 153 AFFOLVGDSESSLYAPVLQPLVGEVTFDFOSARKEISQLSKMLTTTVAQERFAI 212
DB 242 -----PVHFSSWGGGDRG-----NPFVTHVTREVL 270
QY 213 QPKYENAKRCVTQDLRL-SALVSTKCH-----SLGSQSTNFGF-----AKSLTRVENAL 261
DB 271 S-RWKAADLYKDINEISELSMTKCNQDTRVLQAGEDEHPYAILKQLTLLSDTKEIL 329
QY 262 -VHLSGILKAPAKATVEQ-----EVAESSVSSEGEPLSHMDT-----K 299
DB 330 DAKINGOKLAVKAPLQSVQELWDLFLFACYQSLRCGMSMIAEGSL---LDTLRVKAFGV 386
QY 300 HIERIPHASEQAQTVSOHLHAGNISLGNLNNMRDLAFHLLREVSDFYFQSPHPISF 359
DB 387 HLVLDRQESTR-----HADVLSEL-----TRYLG-----IGDYNHSE-QDKIAF 427
QY 360 LLEKAIRGWYLSPELLREMM--SEQNGDALST 390
DB 428 LTNELAS-----KRPLLPRDQWQPSQVKEVLDT 455

RESULT 5
CC37 YEAST STANDARD; PRT; 506 AA.
AC P06101; Q04132;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hsp90 co-chaperone Cdc37 (Hsp90 chaperone protein kinase-targeting
DE subunit) (Cell division control protein 37).
GN CDC37 OR SMO1 OR YDR168W OR YD9489.03.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE OF 58-506 FROM N.A.
RX MEDLINE=86312926; PubMed=3018576;
RA Ferguson J., Ho J.-Y., Peterson T.A., Reed S.I.;
RT "Nucleotide sequence of the yeast cell division cycle start genes
RT CD28 CDC36 CDC37 and CDC39, and a structural analysis of the
RT predicted products.";
RL Nucleic Acids Res. 14:6681-6697 (1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Oliver K., Harris D., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-1995) to the ENBL/GenBank/DBJ databases.

```

---

```

RN RP INTERACTION WITH CDC28.
RX MEDLINE=95273421; PubMed=7753858;
RA Gerber M.R., Farrell A., Deshaies R.J., Herskowitz I., Morgan D.O.;
RT "Cdc37 is required for association of the protein kinase Cdc28 with G1
and mitotic cyclins.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:4651-4655 (1995).
RN [4]
RP INTERACTION WITH STE11.
RX MEDLINE=20130028; PubMed=10664467;
RA Abbas-Terki T., Donze O., Picard D.;
RT "The molecular chaperone Cdc37 is required for Ste11 function and
pheromone-induced cell cycle arrest.";
RL FEBS Lett. 467:111-116 (2000).
RN [5]
RP INTERACTION WITH CDC28 AND CAK1.
RX MEDLINE=20094952; PubMed=10629030;
RA Farrell A., Morgan D.O.;
RT "Cdc37 promotes the stability of protein kinases Cdc28 and Cak1.";
RL Mol. Biol. 20:749-754 (2000).
CC -!- FUNCTION: Co-chaperone that binds to numerous kinases
and promotes their interaction with the Hsp90 complex,
resulting in stabilization and promotion of their activity.
CC -!- SUBUNIT: Forms a complex with Hsp90. Interacts with CDC28, CAK1
and STE11.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the CDC37 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; X04288; CAA27836.1; -.
DR EMBL; 247813; CAA87799.1; -.
DR PIR; S50914; S50914.
DR Germonline; 140659; -.
DR SGD; S0002575; CDC37.
DR GO; GO:0003754; F:chaperone activity; IDA.
DR GO; GO:0000074; P:regulation of cell cycle; IMP.
DR GO; GO:0007103; P:spindle pole body duplication (sensu Saccha. .; IGI.
DR InterPro; IPR004918; Cdc37.
DR Pfam; PF03234; Cdc37; 1.
DR Chaperone; Cell division; Cell cycle.
FT CONFLICT 169 169 AA; 58385 MW; 0DF0C923158A2526 CRC64;
SQ SEQUENCE 506 AA; 58385 MW; 0DF0C923158A2526 CRC64;

Query Match 5.4%; Score 126; DB 1; Length 506;
Best Local Similarity 19.1%; Pred. No. 1.1;
Matches 90; Conservative 71; Mismatches 161; Indels 148; Gaps 17;

QY 15 LSDDSGICGYLKLKSAFRLNEFNVAQTALRKLSQPSADERDALQEAELNWKILSD 74
DB 13 LSDDSDVEVHPNVKKSF-----IKWK--QQ 36
QY 75 SLVEQFSKTRTDIELISWFAAQFLDITLLESAANSLEWLAADLSEKHWDHNPVLPVETL 134
DB 37 SIHEQFRFRNDIKNLETQVDVMSHLNKRVDRI-----LSNLPESSITDLPVTKFLNA 90
QY 135 KSDDDKGKERQADAKVAFQVLGVDSSESSILYAPVLQPLVGEVTFDFQSERKEGI 194
DB 91 NFDKMEKSGENVDPDEIATYNMVEDLFE-----QLA-----KDLDEKGD 131
QY 195 SQLKSMLTITVAQERFAIQKRWENAKRCVTQDLRLSALVSTKCHSLGSQSTNFGFAKSL 254
DB 132 SKSPSLIRDAIKHRAKIDSVTVEAKK-----KLDELKKEKNAHISSEDIHTGDSFM 185
QY 255 TRVENALVHLSGI---KLAPKAEAKTVEQEAESVSE-----GEL 292
DB 186 NKQKGAKPLEATPSEALSSAAES-NILNKLAKSSVPQTFIDFKDDPMKLAKETEFGKI 244

```

```

QY 293 PSHMDTKH-----IERIPNASEQ-----AQTVSQHLHAG----- 321
DB 245 SINEYSKQKQLLEHPLPIISQKDALMKKAFYQLHGDDKRWTLQVHQSELMAVYIKIY 304
QY 322 -----NLSELGNLNMNRDLAFHLRLREVSDFRQSEPHSPISFLLEKAIKRWGYSLLP 373
DB 305 DMKIKIPYLPWELSNVINM-----PFKEV--IFNKDKPMGKESFL--RSVQEKFLHIQ 353
QY 374 E-----LJREMMEQNGDALSTIFNAGLNHLQVLLPEVSTPTVGIESPQ 419
DB 354 KRSLKIQOEEMDESNAEGVETI-----QLKSLDDSTELVNLFPDNFSKDPE 399

RESULT 6
REGA_DICDI
ID REGA_DICDI STANDARD; PRT; 793 AA.
AC Q23917;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUN-1998 (Rel. 36, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 3',5'-cyclic-nucleotide phosphodiesterase rega (EC 3.1.4.17) (PDEase
DE rega).
DE GN REGA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RN STRAIN=AX4;
RP SEQUENCE FROM N.A.
RX MEDLINE=97140317; PubMed=8986798;
RA Shaullsky G., Escalante R., Loomis W.F.;
RT "Developmental signal transduction pathways uncovered by genetic
RT suppressors.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:15260-15265 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AX2;
RA Thomason P.A., Traynor D., Cavet G., Chang W.T., Harwood A.J.,
RA Kay R.R.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Morphological suppressor of tagB.
CC -!- CATALYTIC ACTIVITY: Nucleoside 3',5'-cyclic phosphate + H(2)O =
CC nucleoside 5'-phosphate.
CC -!- DEVELOPMENTAL STAGE: Expressed at low levels in vegetative cells
CC and at high levels in prespore and prestalk cells during
CC development.
CC -!- SIMILARITY: Belongs to the cyclic nucleotide phosphodiesterase
CC family.
CC -!- SIMILARITY: Contains 1 response regulatory domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U60170; AA03508.1; .
CC EMBL; AJ005398; CA006513.1; .
CC DictyBase; DDB0001583; rega.
CC InterPro; IPR003607; Met_Phsphohydro.
CC InterPro; IPR002073; PDEase.
CC InterPro; IPR001789; Response_reg.
CC Pfam; PF00233; PDEase; 1.
CC Pfam; PF00072; response_reg; 1.
CC PRINTS; PR00387; PDIESTERASE1.
CC P-coDom; PD000039; Response_reg; 1.
CC SMART; SM00471; HDC; 1.
CC SMART; SM00448; REC; 1.
CC PROSITE; PS00126; PDEASE_I; 1.
CC PROSITE; PSS0110; RESPONSE REGULATORY; 1.

```

```

KW Hydrolase; CAMP; CGMP.
FT DOMAIN 161 280 RESPONSE REGULATORY.
FT DOMAIN 18 28 POLY-SER.
FT DOMAIN 52 69 POLY-ASN.
FT DOMAIN 87 96 POLY-THR.
FT DOMAIN 100 121 POLY-ASN.
FT DOMAIN 166 169 POLY-ASP.
FT DOMAIN 770 776 POLY-SER.
SQ SEQUENCE 793 AA; 91175 MW; 6E065A96620F8C27 CRC64;

Query Match 5.3%; Score 125; DB 1; Length 793;
Best Local Similarity 19.0%; Pred. NO. 2.2;
Matches 107; Conservative 88; Mismatches 183; Indels 184; Gaps 27;

QY 1 MPLSKHQIEQLSKPLS-----DDSIQGVYIKLE-----KSAFRPLRNEP----- 39
DB 310 MELKEHEIBELTKVSKMSISKEAMESPLVSVTVNIEELLKQSSWSHVESIKEKLSI 369
QY 40 -----NVAQTALRKLSQNSADBERDALQEAQCLNKKILSDSLYEQSKTKTRDIELISW 92
DB 370 LKELGSSNIYRPSFEKLIKNDSDV-----PVTKSLYSEFSST----- 408
QY 93 FVAQAFLDITLESAAANSLEWLADLSEKHWDHLNPVLPVETLKSDDD-----KQKEREQADA 149
DB 409 -----SRNSI-----PTFPQTYNRDTKEVIKWE----- 434
QY 150 KVKARFQVLYGDSSESSILYAPVLQPLVGEVTFDFQSAE-RKGISQLSKSLTTTVAQE 208
DB 435 -----FDVFKYSEDD-----LMLLVDM-FENFQLPETFKIEKLQRIQIMTVNLY 480
QY 209 R-----FAIQFKME-----NAKRCVTQDLRLSALVSTKQHSLSGOSTNFGPAK 251
DB 481 RKNRYHNTHAFDVTQTVYVLTSTFNAAQYLTHLDIFALLISCMCHDLNHFQNTFOV 540
QY 252 SLLTR-----VENALVHLSGIKLAPKAEKTV-----QEVAESVS-----EGE 291
DB 541 NAQTELSLEYNDISVLENHAWLT-FKILRNECNILGNLQEDQKURRVSQVUILAYD 599
QY 292 LPSHMD-----TKHIERIPMASEQAQTVSQHLHAGNLSELGNLNMNRDLAFHL-----L 341
DB 600 MQNHFHTNKKFQHLNLLNLPF--DRNKEDQMLNLFLLIKGDISNIAR--PRLNFEWLSL 655
QY 342 REVSDYFRQSEPHS-----PISFLLEKA-----IRWGL-----SLPELLREMM-- 380
DB 656 RVSDFFQOSHVETICGYPVTFPMDKTKTRARIADFDVFASPLFQSMAPLKESQFL 715
QY 381 -----SEONGDALSTIFNAGLNHLQVLLPEVSTPTVGIESTPQAKPSVSPRSVE 434
DB 716 LKVISKNRNWQAYMELQKEGKCNDD--LQFMEDPILVKS-KLPKIDEEENRDKVSS 771
QY 435 EHVSTSPVDTSKQDQKPSQSS 456
DB 772 SSSSTAPLTSTSSNNETSSS 793

RESULT 7
SMC3_SCHPO
ID SMC3_SCHPO STANDARD; PRT; 1194 AA.
AC O42649;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Structural maintenance of chromosome 3 (Cohesin complex Psm3 subunit).
GN PSM3 OR SMC3 OR SPAC10F6.09C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]_TaxID=4896;
RN SEQUENCE FROM N.A.
RP STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;

```

Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgroes J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald D., Odell C.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., O'Neill J.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Rhode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,  
 RA "The genome sequence of Schizosaccharomyces pombe."  
 RT Nature 415:871-880(2002).  
 RN [2]  
 RN FUNCTION, AND SUBCELLULAR LOCATION AND IDENTIFICATION IN A COHESIN  
 RP COMPLEX WITH PSM1 AND RAD21.  
 RX MEDLINE=20523814; PubMed=11069892;  
 RA Tomonaga T., Nagao K., Kawasaki Y., Furuya K., Murakami A.,  
 RA Morishita J., Yuasa T., Sucani T., Kearsey S.E., Uhlmann F.,  
 RA Namyth K., Yanagida M.;  
 RA "Characterization of fission yeast cohesin: essential anaphase  
 RT proteolysis of Rad21 phosphorylated in the S phase."  
 RL Genes Dev. 14:2757-2770(2000).  
 CC -!- FUNCTION: Involved in chromosome cohesion during cell cycle and in  
 CC DNA repair. Central component of cohesin complex. The cohesin  
 CC complex is required for the cohesion of sister chromatids after  
 CC DNA replication. The cohesin complex apparently forms a large  
 CC proteinaceous ring within which sister chromatids can be trapped.  
 CC At anaphase, the complex is cleaved and dissociates from  
 CC chromatid, allowing sister chromatids to segregate.  
 CC -!- SUBUNIT: Cohesin complexes are composed of the psm1/smci and  
 CC psm3/smcc heterodimer attached via their hinge domain, rad21/sccl  
 CC which link them, and psc3/sccl, which interacts with rad21.  
 CC -!- SUBCELLULAR LOCATION: Nuclear protein. Associates with chromatin.  
 CC Before prophase it is scattered along chromosome arms. At  
 CC anaphase, the rad21 subunit of the cohesin complex is cleaved,  
 CC leading to the dissociation of the complex from chromosomes,  
 CC allowing chromosome separation.  
 CC -!- DOMAIN: The flexible hinge domain, which separates the large  
 CC intramolecular coiled coil regions, allows the heterotypic  
 CC interaction with the corresponding domain of psm1, forming a V-  
 CC shaped heterodimer. The two heads of the heterodimer are then  
 CC connected by different ends of the cleavable rad21 protein,  
 CC forming a ring structure (by similarity).  
 CC -!- SIMILARITY: Belongs to the SMC family. SMC3 subfamily.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL; AL009197; CAA15722.1;  
 CC F01; T37503;  
 CC GenBank; SPombe; SPAC10F6.09c;  
 CC GO; GO:0005694; C:chromosome; IDA.  
 CC GO; GO:0008278; C:cohesin complex; IDA.  
 CC GO; GO:0007064; P:mitotic sister chromatid cohesion; IMP.

DR InterPro; IPR003439; ABC\_transporter.  
 DR InterPro; IPR003405; SMC\_C.  
 DR InterPro; IPR003395; SMC\_C.  
 DR Pfam; PF02483; SMC\_C; 1.  
 DR Pfam; PF02463; SMC\_N; 1.  
 KW Mitosis; Cell cycle; Chromosome partition; ATP-binding; Coiled coil;  
 KW Nuclear protein.  
 FT NP\_BIND 32 39 ATP (POTENTIAL).  
 FT DOMAIN 176 497 COILED COIL (POTENTIAL).  
 FT DOMAIN 498 667 FLEXIBLE HINGE.  
 FT DOMAIN 668 908 COILED COIL (POTENTIAL).  
 FT DOMAIN 970 1006 COILED COIL (POTENTIAL).  
 FT DOMAIN 1097 1132 ALA/ASP-RICH (DA-BOX).  
 FT SEQUENCE 1194 AA; 136850 MW; FD4D1E81A9E2E423 CRC64;  
 SQ  
 Query Match 5.3%; Score 125; DB 1; Length 1194;  
 Best Local Similarity 21.2%; Pred. No. 3.9; Indels 152; Gaps 26;  
 Matches 113; Conservative 84; Mismatches 84;  
 QY 8 IFQLSKPLSDSDSICGVWLKLEKS-----AFRLRN-----EFNVAQTA 45  
 Db 627 ITRAQYARSHOLNGITLSGDSKKGALTAGYRDYRNSRLDAIKNVKTYQIKESDLQES 686  
 QY 46 LRKL-SQNPASDER-----DALQEAACLNKKWLSD-----SLYEQFSKT 83  
 Db 687 LEKCRSEIESFDQKITACLDLQKAQLSKQFERDHIFPKDELVTITGETDLOESMHK 746  
 QY 84 TRDIELISFWVAQAQFLDIT-TLESAAANSLEWLADLSEKHWDHNLNPLVETLKS----- 136  
 Db 747 SRMLELV-----VLEHLELQOANDLK--SELSSSE-MDELDP-KDVEALKSLSGQIE 794  
 QY 137 -----DDDKGKEREQADAKVAFQVGVDSSESSILYAPVQLPLVGEVTFPDPQSAERK 191  
 Db 795 NLSHEFDALIKERAHIEARKTAL-----EYELTNLY--LRNPLKABI-----GSDNR 841  
 QY 192 GEISQKSLMTTVAQERFAIQKVENAKRCVTDRLSALVSTKCHSLGSGSTNFGPAK 251  
 Db 842 IDESELNSVKRSLKYEN-KLOIIKSSSGGLEEQWRINSEISDRKNELESLEE---LQH 897  
 QY 252 SLLTRVENALVHLSGIKLAPKAEAKTVEQEAESSV-----SEGELPSHMDT 298  
 Db 898 EVATRIEQ-----DAXINERNAKRSLLARKECKNEKIKSLGVLPPEEAFI 943  
 QY 299 KRIERIPMASEAQAQTVSHLHAGN--LSELGNLN-----NNNRCLAFHLLREYSD 346  
 Db 944 KYV-----STGSNAIVKHLKHNALKDYGVSNNKAYEQFNNFTKQKD---SLLARREE 994  
 QY 347 YFRQSPHSPISFLL-----EKAIRMGYLSLPELLREMMSEQNGDALSTIFNAAGLN 398  
 Db 995 LRQSQESISELTVLDQKDEALERTFKQVAKSFSEIFVKLVLPAGRGE--LVMNRR--S 1049  
 QY 399 HLDQVLLPVSFTVGVIESPQTPQAKPSVSDSPRSVEEHVSQTSFVDTQSKDQ 451  
 Db 1050 ELSQSIEQDISMDI-----DTPSQKSSIDNYTGIRSVFSNSKODEQLNINO 1096  
 RESULT 8  
 ID MYH3 HUMAN STANDARD; PRT; 1940 AA.  
 AC P11055; Q15492;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Myosin heavy chain, fast skeletal muscle, embryonic (Muscle embryonic  
 DE myosin heavy chain) (SMHCE).  
 GN MYH3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89263803; PubMed=2726495;

RA Eller M.S., Stedman H.H., Sylvester J.E., Fertels S.H.,  
 RA Rubinstein N.A., Kelly A.M., Sarkar S.,  
 RT "Nucleotide sequence of full length human embryonic myosin heavy  
 RT chain cDNA.",  
 RL Nucleic Acids Res. 17:3591-3592(1989).  
 RN [2]  
 RP SEQUENCE OF 774-1940 FROM N.A.  
 RX MEDLINE=90033298; PubMed=2806546;  
 RA Eller M.S., Stedman H.H., Sylvester J.E., Fertels S.H., Wu Q.-L.,  
 RA Raychowdhury M.X., Rubinstein N.A., Kelly A.M., Sarkar S.,  
 RT "Human embryonic myosin heavy chain cDNA. Interspecies sequence  
 RT conservation of the myosin rod, chromosomal locus and isoform  
 RT specific transcription of the gene.",  
 RL FEBS Lett. 256:21-28(1989).  
 RN [3]  
 RP SEQUENCE OF 856-1940 FROM N.A.  
 RC TISSUE=Skeletal muscle;  
 RX MEDLINE=90235862; PubMed=1691980;  
 RA Bober E., Buchberger-Seidl A., Braun T., Singh S., Goedde H.W.,  
 RA Arnold H.H.;  
 RT "Identification of three developmentally controlled isoforms of human  
 RT myosin heavy chains.",  
 RL Eur. J. Biochem. 189:55-65(1990).  
 RN [4]  
 RP SEQUENCE OF 856-1940 FROM N.A.  
 RX MEDLINE=93366648; PubMed=2771643;  
 RA Karsch-Mizrachi I., Travis M., Blau H., Leinwand L.A.;  
 RT "Expression and DNA sequence analysis of a human embryonic skeletal  
 RT muscle myosin heavy chain gene.",  
 RL Nucleic Acids Res. 17:6167-6179(1989).  
 CC -1- FUNCTION: Muscle contraction.  
 CC -1- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2  
 CC heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)  
 CC and 2 regulatory light chain subunits (MLC-2).  
 CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.  
 CC -1- DEVELOPMENTAL STAGE: Abundantly present in fetal skeletal muscle  
 CC and not present or barely detectable in heart and adult skeletal  
 CC muscle.  
 CC -1- DOMAIN: The rodlike tail sequence is highly repetitive, showing  
 CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,  
 CC characteristic for alpha-helical coiled coils.  
 CC -1- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light  
 CC meromyosin (LMW) and 1 heavy meromyosin (HMM). It can later be  
 CC split further into 2 globular subfragments (S1) and 1 rod-shaped  
 CC subfragment (S2).  
 CC -1- SIMILARITY: Contains 1 myosin-like globular head domain.  
 CC -1- SIMILARITY: Contains 1 IQ domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X13988; CAA32167.1; -  
 DR EMBL; X13100; CAA31492.1; -  
 DR EMBL; X51593; CAA35942.1; -  
 DR EMBL; X15696; CAA33731.1; -  
 DR FIR; S04090; S04090.  
 DR HSP; P13538; 2MYS.  
 DR Genew; HGNC:7573; MYH3.  
 DR MIM; 160720; -  
 DR GO; GO:0007517; P:muscle development; TAS.  
 DR InterPro; IPR000048; IQ\_region.  
 DR InterPro; IPR001609; Myosin\_head.  
 DR InterPro; IPR004009; Myosin\_N.  
 DR InterPro; IPR002928; Myosin\_tail.  
 DR Pfam; PF00612; IQ; 2.  
 DR Pfam; PF00063; myosin\_head; 1.  
 DR Pfam; PF02736; Myosin\_N; 1.  
 DR Pfam; PF01576; Myosin\_tail; 1.

DR PRINTS; P00193; MYOSINHEAVY.  
 DR ProDom; PD000355; myosin\_head; 1.  
 DR SMART; SM00015; IQ; 1.  
 DR SMART; SM00242; MYSC; 1.  
 DR PROSITE; PS00096; IQ; 1.  
 KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;  
 KW Calmodulin-binding; ATP-binding; Methylation; Multigene family.  
 FT DOMAIN 1 781  
 FT MYOSIN HEAD-LIKE.  
 FT DOMAIN 782 811  
 FT COILED COIL (POTENTIAL).  
 FT NP\_BIND 840 1933  
 FT ATP (POTENTIAL).  
 FT DOMAIN 179 186  
 FT ACTIN-BINDING.  
 FT DOMAIN 656 678  
 FT ACTIN-BINDING.  
 FT DOMAIN 758 772  
 FT METHYLATION (TRI-) (POTENTIAL).  
 FT MOD\_RES 130 130  
 FT CONFLICT 1331 1331  
 FT A -> G (IN REF. 3).  
 FT CONFLICT 1391 1392  
 FT KK -> OE (IN REF. 1 AND 2).  
 FT CONFLICT 1608 1609  
 FT SR -> RA (IN REF. 3).  
 FT CONFLICT 1663 1664  
 FT RG -> QT (IN REF. 2).  
 SQ SEQUENCE 1940 AA; 224035 MW; 43CA58C6A4BA1253 CRC64;  
 Query Match 5.3%; Score 124.5; DB 1; Length 1940;  
 Best Local Similarity 20.6%; Pred. No. 8;  
 Matches 100; Conservative 86; Mismatches 200; Indels 99; Gaps 20;  
 QY 7 QTEQLSKPLSDSDSICGVYIKLEKSAFR----PLRNFVNAQTALRKLQNPSADERDALQ 62  
 DB 919 KIKEVTERAEDEEINAEITAKRKLEDECSSELKDDIDDELTAKVEKEKATE----- 973  
 QY 63 EACLKWKILSDSLV--EQPSKTRDTIELISFWVAAQFLID--TTLSAANSLEWLADL 117  
 DB 974 ----NKNVNLFEELSGLDETTAKLTREKALQ--EAKQALDDLOAEDKVNLSNKTSK 1027  
 QY 118 SKHWDHNLNPLPVE-TLKSDDDKGRERQADAKV--KAFFQVGDSESSSILYAPVLQL 174  
 DB 1028 LSEQVEDDLESSLEQEKRLVDLERNKRKLEGLKLAQESILDLENDKQ----- 1076  
 QY 175 PLVGEVTFDFQSAERKGEISQLKSMLTITVAQERFAIQ--KMNNAKRCVTQL--DRLSA 231  
 DB 1077 -----LDEKLKKKDFEYQLOQSKVED---EQTGLQFOKKIKELQARIEELSEEIEA 1125  
 QY 232 LVSTKCHSIGSOSTNFGFAKSLITRVENA-LVHLSGIKLAPKAEAK--TVQVEAVESVS 288  
 DB 1126 ERATRAKTEKQSDYARELEELSERLEPAGGVTSQIELNKKRAEFLKLARDLEATLQ 1185  
 QY 289 EGELPSHMDTKHIERIPMASQAQVTSQHLHAGNLSLGNLNNMRDL-----APHILRE 343  
 DB 1186 HEAMVATLRKHADSVAEELGEQ-----IDNLQKVQKLEKSEPKL--E 1228  
 QY 344 VSDYFRQSEPHSPIFLLEKALRWGYLSLPPELLREMMGEQGDALSTIFNAAGLNHLDOV 403  
 DB 1229 IDDLSSSVESVSKKANLEKICR-----TLEDQLSE-----ARGKNEEIQR 1269  
 QY 404 LLPEVST-----PTVGIESPTQPAKPSVD-PRSVEEHVSVQTSVPVTSQKQDKPOSS 456  
 DB 1270 SLSELTTQKSLQTEAGELSRQLEEKESIVSQLSRKQAFTQQTBELKROLEENKAKNA 1329  
 QY 457 ATSAAL 461  
 DB 1330 LAHAL 1334  
 RESULT 9  
 MYH3 RAT  
 ID MYH3 RAT STANDARD; PRT; 1940 AA.  
 AC P12847;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE MYosin heavy chain, fast skeletal muscle, embryonic.  
 GN MYH3.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

NCBI\_TaxID=10116;  
 [1]  
 SEQUENCE FROM N.A.  
 MEDLINE=87060986; PubMed=3783701;  
 RA Strehler E.E., Strehler-Page M.-A., Perriard J.C., Periasamy M.,  
 RA Nadai-Ginard B.;  
 RA "Complete nucleotide and encoded amino acid sequence of a mammalian  
 RT myosin heavy chain gene. Evidence against intron-dependent evolution  
 of the rod.";  
 RL J. Mol. Biol. 190:291-317(1986).  
 CC  
 CC -!- FUNCTION: Muscle contraction.  
 CC -!- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2  
 CC heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)  
 CC and 2 regulatory light chain subunits (MLC-2).  
 CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.  
 CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing  
 CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,  
 CC characteristic for alpha-helical coiled coils.  
 CC -!- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light  
 CC meromyosin (lmm) and 1 heavy meromyosin (hmm). It can later be  
 CC split further into 2 globular subfragments (S1) and 1 rod-shaped  
 CC subfragment (S2).  
 CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.  
 CC -!- SIMILARITY: Contains 1 IQ domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X04267; CAA27817.1; -;  
 CC F1R; A24922; A24922.  
 CC HSP; P13538; 2MYS.  
 CC InterPro: IPR000048; IQ region.  
 CC InterPro: IPR001509; myosin head.  
 CC InterPro: IPR004009; Myosin\_N.  
 CC InterPro: IPR002928; Myosin\_tail.  
 CC InterPro: IPR002017; Spectrin.  
 CC Pfam; PF00612; IQ; 2.  
 CC Pfam; PF00063; myosin head; 1.  
 CC Pfam; PF02736; myosin\_N; 1.  
 CC Pfam; PF01576; Myosin\_tail; 1.  
 CC PRINTS; PR00193; MYOSINHEAVY.  
 CC ProDom; PD000355; myosin\_head; 1.  
 CC SMART; SM00045; IQ; 1.  
 CC SMART; SM00242; MYSC; 1.  
 CC PROSITE; PS00096; IQ; 1.  
 CC Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;  
 KW Calmodulin-binding; ATP-binding; Methylation; Multigene family.  
 FT DOMAIN 1 781  
 FT MYOSIN HEAD-LIKE.  
 FT DOMAIN 782 811  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 840 1933  
 FT NP\_BIND 179 186  
 FT ATP (POTENTIAL).  
 FT DOMAIN 656 678  
 FT ACTIN-BINDING.  
 FT DOMAIN 758 772  
 FT ACTIN-BINDING.  
 FT MOD\_RES 130 130  
 FT METHYLATION (TRI-) (POTENTIAL).  
 SQ SEQUENCE 1940 AA; 223857 MW; B5D546A596E5A696 CRC64;  
 Query Match  
 Best Local Similarity 20.9%; Pred. No. 8;  
 Matches 101; Conservative 90; Mismatches 194; Indels 99; Gaps 21;  
 QY 7 QTEQLSKPLSDSDSGVGVKLEKSAFR---PLRNEFVAOTAKRLKSONPSADERDALQ 62  
 Db 919 KIKVEFRAEDDEEENAEIATKAKRLKDECELSKXDDIDELTLAKVEKEKHATE----- 973  
 QY 63 EACLNKWILSDSLY---EQFSKTTTRDIELISWFAAFLD--TTLESAAANSLEWLADL 117  
 Db 974 ----NKVKNLTBEAGLDETIAKLTRKKALQ--EAAQOTLDDLOAEEDKYNLSLKLKSK 1027

QY 118 SEKHWDHNLNVLVPE-TLKSDDDKKEREQADAKV--KAPFQLYVDSESSILYAPVLQL 174  
 Db 1028 LEQQVDDLESLEQEKRLRVLENRKRLKLEGDLKLAQESILDLENDKQQ----- 1076  
 QY 175 PLVGEVTFDFQSAERKEIGISQLKMLTTVAQER---FAIQFMENAKRCVTQL--DRLS 230  
 Db 1077 -----LDELKXKDFEYQLQS---KYVEDQTLSQLQKKIKELQARIELEEBIE 1124  
 QY 231 ALVSTKCHSLGQSQTNFGFAKSLTRVENA-LVHLSGIKLAPKABAK--TVEQEVASSV 287  
 Db 1125 ABRATRAKTRKQSDYARELEBELSERLEAGGVTSQIELNKKREAEFLKLRDLLEATL 1184  
 QY 288 SBGELPSHMDTKHIERIPMAEQEQAQTVSOHLHAGNLSLGNLNNRDL-----AFHLR 342  
 Db 1185 QHEATVATLRKHAD---SAAELAE-----QIDNLQVKKLEKESEFKL-- 1227  
 QY 343 EYSDYFROSEPHSPISFLLEXAIRMGYLSLPPELLREMMSEONGDALSTIFNAAGLNHLQ 402  
 Db 1228 EIDDLSSSVESVSKANLEKICR-----TLEDQLSE-----ARGKNEETQ 1268  
 QY 403 VLLPEVSTPTVGIESPQTPOAKPSVSDPRSVBEHVSTSPVDTSKQDKPQ---SSAT 458  
 Db 1269 RSLSELTQKSLQT-EAGELSRQLEEKESIVSQLSRSKQAFTQOIELKRLKEENKAK 1327  
 QY 459 SALS 462  
 Db 1328 NALA 1331  
 RESULT 10  
 MYSIP SCHMA  
 ID MYSP SCHMA STANDARD; PRT; 866 AA.  
 AC P06198;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Paramyosin.  
 OS Schistosoma mansoni (Blood fluke).  
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;  
 OC Schistosomatidae; Schistosomatidae; Schistosoma.  
 OC NCBI\_TaxID=6183;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91270282; PubMed=2052029;  
 RA Lactette J.P., Landa A., Arcos L., Willms K., Davis A.E.,  
 RA Shoemaker C.B.;  
 RT "Paramyosin is the Schistosoma mansoni (Trematoda) homologue of  
 RT antigen B from Taenia solium (Cestoda).";  
 RL Mol. Biochem. Parasitol. 44:287-296(1991).  
 RN [2]  
 RP SEQUENCE OF 303-742 FROM N.A.  
 RX MEDLINE=87018840; PubMed=3094144;  
 RA Lanar D.E., Pearce E.J., James S.L., Sher A.;  
 RT "Identification of paramyosin as schistosome antigen recognized by  
 RT intradermally vaccinated mice.";  
 RL Science 234:593-596(1986).  
 CC -!- FUNCTION: Paramyosin is a major structural component of many thick  
 CC filaments isolated from invertebrate muscles.  
 CC -!- SUBUNIT: Homodimer (by similarity).  
 CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.  
 CC -!- SIMILARITY: HIGH, TO MYOSIN HEAVY CHAINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M35499; AAA29915.1; -;  
 CC EMBL; M4163; AAA29914.1; -;  
 DR InterPro; IPR002928; Myosin\_tail.



```

DR Pfam; PF01576; Myosin_tail; 1.
KW Coiled coil; Muscle protein; Thick filament; Myosin.
FT DOMAIN 1 22 NONHELICAL REGION (POTENTIAL).
FT DOMAIN 23 839 COILED COIL (POTENTIAL).
FT DOMAIN 840 866 NONHELICAL REGION (POTENTIAL).
FT DISULFID 750 750 INTERCHAIN (POTENTIAL).
FT CONFLICT 423 428 DOVKEL--> GSSQRI (IN REF. 2).
FT CONFLICT 430 431 SS -> KL (IN REF. 2).
FT CONFLICT 637 637 V -> L (IN REF. 2).
FT CONFLICT 639 639 T -> I (IN REF. 2).
FT CONFLICT 691 691 G -> E (IN REF. 2).
FT CONFLICT 720 720 S -> F (IN REF. 2).
SQ SEQUENCE 866 AA; 100387 MW; 42FA5E6E78176AE0 CRC64;

Query Match 5.3%; Score 124; DB 1; Length 866;
Best Local Similarity 21.3%; Pred. No. 2.9;
Matches 103; Conservative 82; Mismatches 192; Indels 116; Gaps 21;

QY 7 QIEQLSKPLSDSDSICGVYLKLEKSAFRLNFEFNAQTALRKLSQNPSPADERDALQACL 66
Db 122 EVENLQKQG-----KAEDKSHLINEVDNVLGOLDGALKAKQSAESKLEGLDSQL 172
QY 67 NKWKILSDSYEFQSK-----TTSDIEL--ISWFVAAQFLDITLLESAANSLEWLAD 116
Db 173 NRLKSLTDDQLQRLTDLNNAKSLTSENPELLHINQDYEAQIL---NYSKAKSSLESQVD 229
QY 117 LSEKHWDLNPNVLPVETLKDSDDKGKEREQADAKVKAFTLVGD-----SESSIL 167
Db 230 -----DKRSLDDEAKNRFNLQQLTS-LQMDYDNLQAKYDESEASNL 273
QY 168 YAPVLQPLVGEVTFDFQSAERKGEISQLSKMLTTVAQERPAIQFKNENAKRCVTOLD 227
Db 274 RS-----QVSKFNADIAALKSKFE--RELMSKTEEFEMKRFPTWR-----ITELE 317
QY 228 RISALVSTKCHSLGQSQTNEGF-AKSLITRIVENALVHLSCIKLAPKAETVEQVQVARS 286
Db 318 DTAERERLKAVALSEKLTKLTLIKQLQSEIES--LSLENSEIIRAKA-----AESL 368
QY 287 VSEGEPLSHMDTKHIERIPMASQAOVTQSOHLHAGNL-SELGNLNN-MNRDLAFHLLREV 344
Db 369 AS--DLQRRYDELTIETVNTITSONSLESENLRKLSLVNDLTDKNLLERE-----NRQM 421
QY 345 SDVFROSEPHSPISFLEKAIKRGYLSLPEL--LREMMEQNGDALSTIFNA-----394
Db 422 NDQVKE-----LKSRLDANRRLTDLALRSQLEARNDLASALHDAEALHMD 471
QY 395 -----AGLNHLQDVLPLPVSPPTVGTIESPOTPOAKPSVSDPSRVEHVQSPTVDTQS 447
Db 472 QKYQAOQALNHLKSEMEQRLERDELESRK-----STRITIEELTVITIMEVKI 524
QY 448 QKD 450
Db 525 KSE 527

RESULT 11
DPOL_RICPR STANDARD; PRT; 867 AA.
AC O05949;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA polymerase I (EC 2.7.7.7) (POL I).
GN POLA OR RP776.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E;
RX MEDLINE=97419517; PubMed=9274032;
RA Andersson J.O., Andersson S.G.E.;

```

```

RT "Genomic rearrangements during evolution of the obligate
RT intracellular parasite Rickettsia prowazekii as inferred from an
RT analysis of 52015 bp nucleotide sequence.";
RN Microbiology 143:2783-2795(1997).
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sichertz-Ponten T., Almark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RL Nature 396:133-140(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E, and B;
RX MEDLINE=99416441; PubMed=10486973;
RA Andersson J.O., Andersson S.G.E.;
RT "Genome degradation is an ongoing process in Rickettsia.";
RL Mol. Biol. Evol. 16:1178-1191(1999).
CC -!- FUNCTION: In addition to polymerase activity, this DNA polymerase
CC exhibits 3' to 5' and 5' to 3' exonuclease activity (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + [DNA] (N).
CC -!- SUBUNIT: Single-chain monomer with multiple functions (By
CC similarity).
CC -!- SIMILARITY: Belongs to the DNA polymerase type-A family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Y11784; CAA72476.1; -.
DR EMBL; AJ235273; CAA15203.1; -.
DR EMBL; AJ238755; CAB56085.1; -.
DR EMBL; AJ238756; CAB56089.1; -.
DR PIR; C71638; C71638.
DR HSP; P00582; 1KFS.
DR InterPro; IPR008918; 5_3_exo_C.
DR InterPro; IPR002421; 5_3_exonuclease.
DR InterPro; IPR001098; DNA_pol.
DR InterPro; IPR002298; DNA_pol.
DR InterPro; IPR000513; Exo_N_1.
DR Pfam; PF02739; 5_3_exonuc_N; 1.
DR Pfam; PF01367; 5_3_exonuclease; 1.
DR Pfam; PF00476; DNA_pol_A; 1.
DR PRINTS; PR00868; DNA_POLI.
DR SMART; SM00475; 53EXOC; 1.
DR SMART; SM00279; HhH2; 1.
DR SMART; SM00482; POLAC; 1.
DR TIGRFAMs; TIGR00593; pola; 1.
DR PROSITE; PS00447; DNA_POLYMERASE_A; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication; DNA repair;
KW Hydrolase; Exonuclease; DNA-binding; Complete proteome.
FT DOMAIN 1 283 EXONUCLEASE
SQ SEQUENCE 867 AA; 98651 MW; 557194B38FB69BA2 CRC64;

Query Match 5.3%; Score 123.5; DB 1; Length 867;
Best Local Similarity 21.2%; Pred. No. 3.1;
Matches 90; Conservative 60; Mismatches 172; Indels 103; Gaps 17;

QY 18 DSICGVYLKLEKSAFRLNFEFNAQTALRKLSQNPSPADERDALQACLNKKWILSDSLY 77
Db 186 DNIQGVPSIGPKTA-SSLITKGSVENIFNLSLDQISSIKQKTKLQNA-----231
QY 78 EQFSKTRDIELISWFAAQFLDIT--LESAANSLEWLADLSEKHWDLNPNVLPVETLK 135

```

Db 232 -----REALISWKLIG-----LDSNVLDLFDLNNLKWSPNSKK-----ITGFLQBYGFK 277  
 Qy 136 SDDKGREREQADAKVKAFFQL-VGDSESSILYAPVLQPLVGEVTFDFQSAERKEI 194  
 Db 278 S-----LYKRVENLFDIKINDHEE-----IVDNKVTEAK-EI 308  
 Qy 195 SOLKSMITTTVAQR-----FALQFKENAKRCVTOLDRLSALVSTKCHSLGQSSTNFG 248  
 Db 309 SNASELANFAKEAERIGIFGYLLQKQGEN-RALILSLQNCQYIIKITNNYINIKNNNDW 367  
 Qy 249 FAKSLLRIVEN-----ALVHLSGIKLAPKAEAKTEQ-EVAESSVSEGEPLPSHMDTK 299  
 Db 368 FSHILNLLTNKSIKITYLSKHLKFKYANQSHOITALEDLMLQVYALSAGLQKLNFTK 427  
 Qy 300 HIERIPMAEQAOQTVSQ-HHAGNLSELNMMNNRDLAFHLLREVSDFRQSEPHSPI 357  
 Db 428 TLTKDNIINESARIVINISLYKQTLLEL-----QKNKAFRLYREID-----LPT 472  
 Qy 358 SFLEKARWGYLSLPELLREMMSEQNDAL-----STIFNAAGLNHLDOVLPE 407  
 Db 473 CFILDKMEKVIKVDANYLNLSDFGTETLKIEEIPFALSGTKFNIGSQKQGLGILFKK 532  
 Qy 408 VSTPT 412  
 Db 533 MOLPS 537  
 RESULT 12  
 MYH7 HUMAN  
 ID MYH7 HUMAN STANDARD; PRT; 1935 AA.  
 AC P12883; Q14904; Q16579;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE MYH7 heavy chain, cardiac muscle beta isoform (MyHC-beta).  
 GN MYH7 OR MYHCB.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91065634; PubMed=2249844;  
 RA Jaenicke T., Diederich K.W., Haas W., Schleich J., Lichter P.,  
 Pfordt M., Bach A., Vosberg H.P.;  
 RT "The complete sequence of the human beta-myosin heavy chain gene and  
 a comparative analysis of its product.";  
 RL Genomics 8:194-206(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90301496; PubMed=2362820;  
 RA Liew C.-C., Sole M.J., Yamauchi-Takihara K., Kellam B.,  
 Anderson D.H., Lin L., Liew J.;  
 RT "Complete sequence and organization of the human cardiac beta-myosin  
 heavy chain gene.";  
 RL Nucleic Acids Res. 18:3647-3651(1990).  
 RN [3]  
 RP SEQUENCE OF 1-115 FROM N.A.  
 RX MEDLINE=89264452; PubMed=2726733;  
 RA Yamauchi-Takihara K., Sole M.J., Liew J., Ing D., Liew C.-C.;  
 RT "Characterization of human cardiac myosin heavy chain genes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:3504-3508(1989).  
 RN [4]  
 RP ERRATUM.  
 RA Yamauchi-Takihara K., Sole M.J., Liew J., Ing D., Liew C.-C.;  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:7416-7417(1989).  
 RN [5]  
 RP SEQUENCE OF 1310-1935 FROM N.A.  
 RX MEDLINE=86176778; PubMed=2421254;  
 RA Saez L., Leinwand L.A.;  
 RT "Characterization of diverse forms of myosin heavy chain expressed in  
 adult human skeletal muscle.";  
 RL Nucleic Acids Res. 14:2951-2969(1986).

RN [6]  
 RP REVISIONS.  
 RA Leinwand L.A.;  
 RL Submitted (MAR-1988) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP SEQUENCE OF 1410-1935 FROM N.A.  
 RX MEDLINE=88299163; PubMed=2969919;  
 RA Kurabayashi M., Tsuchimochi H., Komuro I., Takaku F., Yazaki Y.;  
 RT "Molecular cloning and characterization of human cardiac alpha- and  
 beta-form myosin heavy chain complementary DNA clones. Regulation of  
 expression during development and pressure overload in human  
 atrium.";  
 RL J. Clin. Invest. 82:524-531(1988).  
 RN [8]  
 RP SEQUENCE OF 785-1935 FROM N.A.  
 RC TISSUE=Skeletal muscle;  
 RX MEDLINE=90235862; PubMed=1691980;  
 RA Bober E., Buchberger-Seidl A., Braun T., Singh S., Goedde H.W.,  
 Arnold H.H.;  
 RT "Identification of three developmentally controlled isoforms of human  
 myosin heavy chains.";  
 RL Eur. J. Biochem. 189:55-65(1990).  
 RN [9]  
 RP SEQUENCE OF 1393-1935 FROM N.A.  
 RX MEDLINE=87192738; PubMed=3032769;  
 RA Jandreski M.A., Liew C.-C.;  
 RT "Construction of a human ventricular cDNA library and  
 characterization of a beta myosin heavy chain cDNA clone.";  
 RL Hum. Genet. 76:47-53(1987).  
 RN [10]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=96039076; PubMed=8533830;  
 RA Arai S., Matsuka R., Hirayama K., Sukurai H., Tamura M., Ozawa T.,  
 Kimura M., Imamura S.-I., Furutani Y., Jch-O K., Kawana M., Takao A.,  
 Hosoda S., Momma K.;  
 RT "Missense mutation of the beta-cardiac myosin heavy-chain gene in  
 hypertrophic cardiomyopathy.";  
 RL Am. J. Med. Genet. 58:267-276(1995).  
 RN [11]  
 RP VARIANTS CMH1 GLU-256 AND ARG-741.  
 RX MEDLINE=93248216; PubMed=8483915;  
 RA Fananapazir L., Dalakas M.C., Cyran F., Cohn G., Epstein N.D.;  
 RT "Missense mutations in the beta-myosin heavy-chain gene cause central  
 core disease in hypertrophic cardiomyopathy.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:3993-3997(1993).  
 RN [12]  
 RP VARIANT CMH1 GLN-403.  
 RX MEDLINE=90367131; PubMed=1975517;  
 RA Geisterfer-Lowrance A.A.T., Kass S., Tanigawa G., Vosberg H.-P.,  
 McKenna W., Seidman C.E., Seidman J.G.;  
 RT "A molecular basis for familial hypertrophic cardiomyopathy: a beta  
 cardiac myosin heavy chain gene missense mutation.";  
 RL Cell 62:999-1006(1990).  
 RN [13]  
 RP VARIANTS CMH1 GLN-249; GLN-403; CYS-453 AND MET-606.  
 RX MEDLINE=92204193; PubMed=1552912;  
 RA Watkins H., Rosenzweig A., Hwang D.S., Levi T., McKenna W.,  
 Seidmann C.E., Seidmann J.G.;  
 RT "Characteristics and prognostic implications of myosin missense  
 mutations in familial hypertrophic cardiomyopathy.";  
 RL New Engl. J. Med. 326:1108-1114(1992).  
 RN [14]  
 RP VARIANTS CMH1 GLN-403; ARG-453; ARG-584 AND MET-606.  
 RX MEDLINE=94070863; PubMed=8250038;  
 RA Watkins H., Thierfelder L., Anan R., Jarcho J., Matsumori A.,  
 McKenna W., Seidman J.G., Seidman C.E.;  
 RT "Independent origin of identical beta cardiac myosin heavy-chain  
 mutations in hypertrophic cardiomyopathy.";  
 RL Am. J. Hum. Genet. 53:1180-1185(1993).  
 RN [15]  
 RP VARIANTS CMH1 GLN-403 AND VAL-908.  
 RX MEDLINE=92346810; PubMed=1638703;  
 RA Epstein N.D., Cohn G.M., Cyran F., Fananapazir L.;

RT "Differences in clinical expression of hypertrophic cardiomyopathy associated with two distinct mutations in the beta-myosin heavy chain gene. A 908Leu-->Val mutation and a 403Arg-->Gln mutation.";

RL Circulation 86:345-352(1992).

RL [16]

RP VARIANTS CMH1 LEU-403 AND TRP-403.

RX MEDLINE=94075629; PubMed=8254035;

RA Dausse E., Komajda M., Fellet L., Dubourg O., Dufour C., Carrier L., Wsnesky C., Bercovic J., Hengstenberg C., Al-Mahdawi S.;

RA "Familial hypertrophic cardiomyopathy. Microsatellite haplotyping and identification of a hot spot for mutations in the beta-myosin heavy chain gene.";

RT J. Clin. Invest. 92:2807-2813(1993).

RL [17]

RP VARIANTS CMH1 TRP-403.

RX MEDLINE=94093568; PubMed=8268932;

RA Moolman J.C., Brink P.A., Corfield V.A.;

RT "Identification of a new missense mutation at Arg403, a CpG mutation hotspot, in exon 13 of the beta-myosin heavy chain gene in hypertrophic cardiomyopathy.";

RL Hum. Mol. Genet. 2:1731-1732(1993).

RL [18]

RP VARIANTS CMH1 ASN-615.

RX MEDLINE=93038688; PubMed=1417858;

RA Nishi H., Kimura A., Harada H., Toshima H., Sasazuki T.;

RT "Novel missense mutation in cardiac beta myosin heavy chain gene found in a Japanese patient with hypertrophic cardiomyopathy.";

RL Biochem. Biophys. Res. Commun. 188:379-387(1992).

RL [19]

RP VARIANTS CMH1 GLY-778.

RX MEDLINE=93343938; PubMed=8343162;

RA Harada H., Kimura A., Nishi H., Sasazuki T., Toshima H.;

RT "A missense mutation of cardiac beta-myosin heavy chain gene linked to familial hypertrophic cardiomyopathy in affected Japanese families.";

RL Biochem. Biophys. Res. Commun. 194:791-798(1993).

RL [20]

RP VARIANTS CMH1 VAL-908.

RX MEDLINE=93168485; PubMed=8435239;

RA Al-Mahdawi S., Chamberlain S., Cleland J., Nihoyannopoulos P., Gilligan D., French J., Choudhury L., Williamson R., Oakley C.;

RT "Identification of a mutation in the beta cardiac myosin heavy chain gene in a family with hypertrophic cardiomyopathy.";

RL Br. Heart J. 69:136-141(1993).

RL [21]

RP VARIANTS CMH1 TRP-719.

RX MEDLINE=95179132; PubMed=7874131;

RA Greve G., Bachinski L., Friedman D.L., Czernuzewicz G., Anan R., Towbin J., Seidman C.E., Roberts R.;

RT "Isolation of a de novo mutant myocardial beta MHC protein in a pedigree with hypertrophic cardiomyopathy.";

RL Hum. Mol. Genet. 3:2073-2075(1994).

RL [22]

RP VARIANTS CMH1 CYS-513; ARG-716 AND TRP-719.

RX MEDLINE=94110336; PubMed=8282799;

RA Anan R., Greve G., Thierfelder L., Watkins H., McKenna W., Solomon S., Vecchio C., Shono H., Nakao S., Tanaka H., Mares A. Jr., Towbin J.A., Spirito P., Roberts R., Seidman J.G., Seidman C.E.;

RT "Prognostic implications of novel beta cardiac myosin heavy chain gene mutations that cause familial hypertrophic cardiomyopathy.";

RL J. Clin. Invest. 93:280-285(1994).

RL [23]

RP VARIANTS CMH1 THR-797.

RX MEDLINE=96047159; PubMed=7581410;

RA Moolman J.C., Brink P.A., Corfield V.A.;

RT "Identification of a novel Ala797Thr mutation in exon 21 of the beta-myosin heavy chain gene in hypertrophic cardiomyopathy.";

RL Hum. Mutat. 6:197-198(1995).

RL [24]

RP VARIANTS CMH1 CYS-453.

RX MEDLINE=9620901; PubMed=8655135;

RA Ko Y.-L., Chen J.-J., Tang T.-K., Cheng J.-J., Lin S.-Y., Liou Y.-C., Kuan P., Wu C.-W., Lien W.-P., Liew C.-C.;

RT "Malignant familial hypertrophic cardiomyopathy in a family with a 453Arg-->Cys mutation in the beta-myosin heavy chain gene: coexistence of sudden death and end-stage heart failure.";

RL Hum. Genet. 97:585-590(1996).

RL [25]

RP VARIANTS CMH1 THR-349 AND TRP-719.

RX MEDLINE=98204402; PubMed=9544842;

RA Jeschke B., Uhl K., Weist B., Schroder D., Meitinger T., Dohlemann C., Vosberg H.-P.;

RA "A high risk phenotype of hypertrophic cardiomyopathy associated with Query Match 5.2%; Score 123; DB 1; Length 1935; Best Local Similarity 22.1%; Pred. No. 9.8; Indels 128; Gaps 19; Matches 94; Conservative 63; Mismatches 140;

QY 7 QIEQLSKPL-----SDSICGVYLK---LEKSAFRLRNEF-----NVAQTALR 47

DB 1395 ELEEAKKLAQLQAEAEAVNAKSSLEKTKHR-LQNEIEDLMVDVRSNAAAALD 1443

QY 48 KLSQN-----PSADERDALQAC-----LNKWKILSD 74

DB 1444 KQRFNFKILAEWKQYEESSQSELESSTELFKLNAYEESLEHLETFKRENK 1503

QY 75 SLYEQFS-----KTRDIELISWFAAQFL-IDTTLESAAANSLEWLADLSEKHW 123

DB 1504 NQEEISDLTQLGSSGKTIHELEKVRQLEAKWELQSALEAEASLE-----HEGSK 1557

QY 124 HUNPVLVETLKSD-----DDKGEREOA---DAKVKAFFQIVGDSESSILYAPVLQ 175

DB 1558 ILRAQLEFNQIKAEIERKLAEKDEMEQAKENHLRVVDSLQTSLDAETRSRNEALRVKK 1617

QY 176 LVGEVTFPFDFO-----SAERKEISOLKSMLTITVAQ-----ERPFI--- 212

DB 1618 MEGDNLNEMEIQLSHANRWAAEAQVKLSQLSKDTQIQDDAVRANDDKENAIIVRR 1677

QY 213 ---QFKMENAKRCVTQLDRLSAL-----VSTKCHSLGSGSTNFGFAK-----SL 253

DB 1678 NNLQAELEELRAVVEQTERSKLAQLAQLIETSERVQLHSGNTSLNQKKMDADLSQL 1737

QY 254 LTRVENALVHLSGILAPKAETVQEVQVAVSSVSEGEGLPSHMDTK-HIERIPMASEQAQ 312

DB 1738 QTEVEEA-----VOECRNABEK-AKKAITDAAMWAEELKKEQDTSALHRRKKMGEQTI 1790

QY 313 TVSQH 317

DB 1791 KDLQH 1795

RESULT 13

MYH7 RAT

ID MYH7 RAT STANDARD; PRT; 1935 AA.

AC P02564;

DT 21-JUL-1986 (Rel. 01, Created)

DT 01-OCT-1989 (Rel. 12, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Myosin heavy chain, cardiac muscle beta isoform (MyHC-beta).

GN MYH7.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Heart;

RX MEDLINE=90016823; PubMed=2798112;

RA Kraft R., Bravo-Zehnder M., Taylor D., Leinward L.A.;

RT "Complete nucleotide sequence of full length cDNA for rat beta cardiac myosin heavy chain.";

RL Nucleic Acids Res. 17:7529-7530(1989).

RL [2]

RP DISCUSSION OF SEQUENCE.

RN MEDLINE=9013919; PubMed=2614840;

RA McNally E.M., Kraft R., Bravo-Zehnder M., Taylor D., Leinward L.A.;

RT "Full-length rat alpha and beta cardiac myosin heavy chain sequences. Comparisons suggest a molecular basis for functional differences.";  
 RL J. Mol. Biol. 210:665-671(1989).  
 RN [3]  
 RP SEQUENCE OF 1524-1935 FROM N.A.  
 RX MEDLINE=82220036; PubMed=7045682;  
 RA Mahdavi V., Perlasamy M., Nadal-Ginard B.;  
 RT "Molecular characterization of two myosin heavy chain genes expressed in the adult heart.";  
 RL Nature 297:659-664(1982).  
 RN [4]  
 RP SEQUENCE OF 1871-1935 FROM N.A.  
 RX STRAIN=Wistar; TISSUE=Heart;  
 RA MEDLINE=85179510; PubMed=6241892;  
 RT "Cardiac myosin heavy chain isozymic transitions during development and under pathological conditions are regulated at the level of mRNA availability.";  
 RL Eur. Heart J. 5:181-191(1984).  
 RN [5]  
 RP FUNCTION: Muscle contraction.  
 CC -!- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2 heavy chain subunits (MHC), 2 alkali light chain subunits (MLC) and 2 regulatory light chain subunits (MLC-2).  
 CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.  
 CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing cycles of a 28-residue repeat pattern composed of 4 heptapeptides, characteristic for alpha-helical coiled coils.  
 CC -!- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light meromyosin (lmm) and 1 heavy meromyosin (hmm). It can later be split further into 2 globular subfragments (S1) and 1 rod-shaped subfragment (S2).  
 CC -!- MISCELLANEOUS: The cardiac alpha isoform is a 'fast' ATPase myosin, while the beta isoform is a 'slow' ATPase.  
 CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.  
 CC -!- SIMILARITY: Contains 1 IQ domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; X15939; CAA34065.1; -;  
 DR EMBL; J00752; AAA41654.1; -;  
 DR EMBL; M32698; AAA41659.1; -;  
 DR PIR; S06006; S06006;  
 DR HSP; P08799; 1KND.  
 DR InterPro; IPR000048; IQ\_region.  
 DR InterPro; IPR001609; myosin\_head.  
 DR InterPro; IPR004009; Myosin\_N.  
 DR InterPro; IPR002928; Myosin\_tail.  
 DR Pfam; PF00612; IQ; 2.  
 DR Pfam; PF00063; myosin\_head; 1.  
 DR Pfam; PF02736; Myosin\_N; 1.  
 DR Pfam; PF01576; Myosin\_tail; 1.  
 DR PRINTS; PR00193; MYOSIN\_HEAVY.  
 DR ProDom; PD000355; myosin\_head; 1.  
 DR SMART; SM00015; IQ; 1.  
 DR SMART; SM00242; MYSC; 1.  
 DR PROSITE; PS00096; IQ; 1.  
 KW Myosin. Muscle protein. Coiled coil; Thick filament; Actin-binding; ATP-binding; Methylation; Multigene family; Calmodulin-binding.  
 FT DOVAIN 1 780 MYOSIN HEAD-LIKE.  
 FT DOVAIN 781 810 IQ.  
 FT DOVAIN 840 1935 COILED COIL (POTENTIAL).  
 FT NP\_BIND 178 185 ATP.  
 FT DOVAIN 655 677 ACTIN-BINDING.  
 FT DOVAIN 757 771 ACTIN-BINDING.  
 FT MOD\_RES 129 129 METHYLATION (TR1-) (POTENTIAL).  
 FT CONFLICT 1529 1531 IRLK -> VRR (IN REF. 3).  
 FT 1731 1731 D -> H (IN REF. 3).

FT CONFLICT 1784 1784 N -> K (IN REF. 3).  
 FT CONFLICT 1851 1851 T -> N (IN REF. 3).  
 FT CONFLICT 1858 1858 R -> K (IN REF. 3).  
 SQ SEQUENCE 1935 AA; 223082 MW; C8376C324A7BD82B CRC64;  
 Query Match 5.2%; Score 123; DB 1; Length 1935;  
 Best Local Similarity 22.1%; Pred. No. 9.8;  
 Matches 102; Conservative 65; Mismatches 154; Indels 140; Gaps 21;  
 QY 7 QIEQLSKPL-----SDSICGVYK---LKSAPRPLRNEF-----NVAQTALR 47  
 DB 1385 ELEEAKKLAQRLOQAEAEAVNAKCSSLEKTKR-LQNEIDMLVDVRSNAAAAALD 1443  
 QY 48 KLSQN-----PSADERDALQEAQ-----LNKWKILSD 74  
 DB 1444 KKQNFDKILVWKQVYESQSELESQKARSLSLSTELFKLKNAYERSLEHLETKRENK 1503  
 QY 75 SIYEQFS-----KTRDIELISWFAAQFL-LDTTLESANSLWDLADLSEKHW 123  
 DB 1504 MLQEISDLTQLGSGTSGKSIHELEKIRKQLEAEKLEQSALEAEASLE-----HEEGK 1557  
 QY 124 HLNPLVPVETLKSD-----DDKGEREQ-----DAKRAFFQVGVGDSSESSILYAPVLQLP 175  
 DB 1558 ILRAQLEFNQIKAEIERKLAEKDEMEQAKRNHLRVVDSLOTSLDAETRSRNEALRVKKK 1617  
 QY 176 LVGEVTFDFQ-----SAERKEISOLKSMITTTVAQ-----ERFAI----- 212  
 DB 1618 MEGDLNEMEIQLSHANRVAAEAQKVKSLQSLKDTQIQDDAVRANDDKKNIAIVERR 1677  
 QY 213 ---QFKVENAKRCVTQDLRLSAL-----VSTKCHSLGSGSTNFGFAK-----SL 253  
 DB 1678 NNLLQAELELRVAVQETRSKLAQEELIETSERVQLLHSONSLNQKKKMDADLSQL 1737  
 QY 254 LTRVENALVHLSGKIKAPAKAKTVEQVBAESSVSEGEPLPSHMDTK-HIERIPNASQAO 312  
 DB 1738 QTEVEEA-----VQECRNABEK-AKKAITDAAMMAEELKEQDTSALHER--MKNNNEQ 1788  
 QY 313 TVSQHLH-----AGNLSELGNLNMNRDLAFHLRE 343  
 DB 1789 TIKDLQRLDEAEQALKGGKQKQLEARVRELENELEAE 1829  
 RESULT 14  
 ANTI ONCVO STANDARD; PRT; 2022 AA.  
 ID ANT: ONCVO  
 AC P21249; (Rel. 18, Created)  
 DT 01-MAY-1991 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 39, Last annotation update)  
 DE Major antigen.  
 GN OVTL  
 OS Onchocerca volvulus.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;  
 OC Onchocercidae; Onchocerca.  
 OX NCBI\_TaxID=6282;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95287898; PubMed=7770081;  
 RA Tricteatrapapab S., Richie T.L., Tuan R.S., Shepley K.J., Dinman J.D.,  
 RA Neubert I.A., Scott A.L.;  
 RT "Molecular cloning of a gene expressed during early embryonic development in Onchocerca volvulus.";  
 RL Mol. Biochem. Parasitol. 69:161-171(1995).  
 RN [2]  
 RP SEQUENCE OF 733-866 FROM N.A.  
 RX MEDLINE=89127417; PubMed=2464764;  
 RA Donelson J.E., Duke B.O.L., Moser D., Zeng W., Erondou N.E.,  
 RA Lucius R., Renz A., Karam M., Flores G.Z.;  
 RT "Construction of Onchocerca volvulus cDNA libraries and partial characterization of the cDNA for a major antigen.";  
 RL Mol. Biochem. Parasitol. 31:241-250(1988).  
 CC -!- FUNCTION: May be a myofibrillar protein.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U12681; AAA80009.1; --  
 DR EMBL; J03995; AAA29412.1; --  
 DR PIR; T43214; T43214; --  
 KW Antigen; Coiled coil.  
 FT DOMAIN 74 120 COILED COIL (POTENTIAL).  
 FT DOMAIN 151 251 COILED COIL (POTENTIAL).  
 FT DOMAIN 327 384 COILED COIL (POTENTIAL).  
 FT DOMAIN 417 1879 COILED COIL (POTENTIAL).  
 SQ SEQUENCE 2022 AA; 237341 MW; B7132AACF1520317 CRC64;

Query Match 5.2%; Score 123; DB 1; Length 2022;  
 Best Local Similarity 22.9%; Pred. No. 10;  
 Matches 85; Conservative 64; Mismatches 140; Indels 82; Gaps 16;

QY 35 LRNEFNAQTALRLKSONPSADERDALQEAQKILKSLVYEQSKTRDTIELISWFV 94  
 Db 971 LRSLKNDARADL-QN-----RDSILRENDWKEKSDALNWLDR-LRD-ELLVSR 1022  
 QY 95 AAFELL---DTLESAA-NSLEWLADLSEKHDPHLPVLPVETLKSDDDKGRERQADAK 150  
 Db 1023 DAEKEINRYNTDLQTAARNEIKLTPTNEMKSQLN-----AAEDK 1063  
 QY 151 VKAFQVLGVSESSILYAPVLQPLNGEVYTFDFQSAERKGRISQLKSLMTTVAQERF 210  
 Db 1064 INSLNKVITQCKNI-----RDLTGEVHHLEGEKLDKAGNVANLESLDTT-----RE 1111  
 QY 211 AIOFQFENAKRCVTLQDLRLSALVSTKCHSLGSGSTNFGFAKSLTLTRVENALVHLSGKLA.270  
 Db 1112 RIHLLEGONASLOTELNKIK-----GDIDSLFG-----ENDM-LKTAKES 1150  
 QY 271 PKAEKTVGEVEAEVSSVSELSHMDTKHRIPIWASQAOQVSHLAGNLSLGNLN 330  
 Db 1151 NEAEIDLKQKQORSIENAKYSDALDKLRPEYDRL-----QNLVREKIKQAEINT 1201  
 QY 331 NMNRDLAFHL-----LREVSDFPROSE-PHSPISFLEKAIKRWGLSLPEILLRWMSQ 383  
 Db 1202 QAVQDLERLNSQERELRDATDKLIASEGDRNALRSEVKEQLQEVQFMREQLLRK--TDE 1259  
 QY 384 NGDALSTIFNA 394  
 Db 1260 YQAALSDLVNA 1270

RESULT 15  
 LBN\_MOUSE STANDARD; PRT; 1220 AA.  
 ID Q8K1G2; Q8BRF3;  
 AC 15-MAR-2004 (Rel. 43, Created)  
 DT 15-MAR-2004 (Rel. 43, Last sequence update)  
 DE 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Limbin.  
 GN EVC2 OR LBN.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID:10090;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.  
 RC TISSUE=Limb bud;  
 RX MEDLINE=22155879; PubMed=12136126;  
 RA Takeda H., Takami M., Oguni T., Tsuji T., Yoneda K., Sato H.,  
 RA Ihara N., Itoh T., Kata S.R., Mishina Y., Womack J.E., Moritomo Y.,  
 RA Sugimoto Y., Kunieda T.  
 RT "Positional cloning of the gene limbin responsible for bovine  
 RT chondrodysplastic dwarfism."

RL Proc. Natl. Acad. Sci. U.S.A. 99:10549-10554 (2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N; TISSUE=Kidney;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L.H., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madden A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [3]  
 RP SEQUENCE OF 21-744 FROM N.A.  
 RC STRAIN=C57BL/6J;  
 RX MEDLINE=22354693; PubMed=12466851;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nikaido I., Otsu N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 RA Baldarelli R., Hill D.P., Malt C., Hume D.A., Quackenbush J.,  
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
 RA Grimmond S., Guselnich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Nunata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,  
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Saitana R., Schneider C., Semple C.A., Setou M., Shimada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573 (2002).  
 CC -!- FUNCTION: Plays a critical role in bone formation and skeletal  
 CC development. May be involved in early embryonic morphogenesis.  
 CC -!- TISSUE SPECIFICITY: Expressed in long and cranial bones, kidney  
 CC and heart. Strongly expressed in proliferating chondrocytes,  
 CC osteoblasts and osteoclasts.  
 CC -!- DEVELOPMENTAL STAGE: Found in the embryo at day E7, E11, E15, and  
 CC E17. At the limb bud formation stage E11, it is expressed in fore-  
 CC and hindlimb buds, branchial arches, and facial primordia.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2004, 02:06:04 ; Search time 72 Seconds  
(without alignments)  
2028.957 Million cell updates/sec

Title: US-10-780-347-2

Perfect score: 2343  
Sequence: 1 MPLSKHQIEQLSKPLSDDSI.....DTQSKQDKFQSQATSALSW 463

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phase:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	291	12.4	377	16	Q8D6T7
2	284.5	12.1	373	16	Q87HC9
3	252	10.8	358	2	Q93ED3
4	241.5	10.3	351	16	Q8U7V6
5	232.5	9.9	337	16	Q8XRS2
6	209	8.9	366	16	Q91IB9
7	174	7.4	361	16	Q88IB1
8	152.5	6.5	1454	4	Q9UPS3
9	152.5	6.5	1503	4	Q8IWU2
10	151	6.4	2245	5	Q86A36
11	143.5	6.1	959	5	Q86TW2
12	143	6.1	884	5	Q02239
13	141.5	6.0	344	16	Q91750
14	141	6.0	1534	5	Q8MPV7
15	141	6.0	1538	5	Q8MPV6
16	140.5	6.0	3742	8	Q9TKW8

17	137	5.8	347	16	Q8PFF73
18	136.5	5.8	735	5	Q9VMT8
19	136	5.8	1216	16	Q8EAA0
20	134.5	5.7	1179	16	Q8DPJ9
21	133.5	5.7	357	16	Q7UL73
22	130	5.5	16223	5	Q8IR22
23	129.5	5.5	630	5	Q25613
24	129.5	5.5	1133	5	Q21022
25	127.5	5.4	2471	12	Q9VTK3
26	127	5.4	1935	6	Q9BE39
27	126.5	5.4	752	4	Q13597
28	126.5	5.4	893	4	Q8K6Z0
29	126.5	5.4	1023	16	Q92B64
30	126.5	5.4	2094	11	Q80Y35
31	126	5.4	613	2	Q9XD53
32	126	5.4	1029	16	Q8IDR0
33	125.5	5.4	1992	13	Q04834
34	125.5	5.4	2167	2	Q845L8
35	125	5.3	1935	6	Q9GKR1
36	124.5	5.3	550	3	Q59725
37	124.5	5.3	1964	13	Q93522
38	124	5.3	1935	6	Q8MJU9
39	123.5	5.3	1051	2	Q49524
40	123	5.2	611	4	Q14905
41	123	5.2	800	10	Q94GM1
42	123	5.2	1935	4	Q9H1D5
43	123	5.2	2471	13	Q9DDN8
44	122.5	5.2	1220	11	Q8K1G2
45	122	5.2	1344	2	Q49545

#### ALIGNMENTS

#### RESULT 1

Q8D6T7 PRELIMINARY; PRT; 377 AA.  
 ID Q8D6T7  
 AC Q8D6T7;  
 DT 01-MAR-2003 (TREMELrel. 23, Created)  
 DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
 DE Uncharacterized protein conserved in bacteria.  
 GN VV20438.  
 OS Vibrio vulnificus.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrio.  
 CX NCBI\_TaxID=672;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CMCP6;  
 RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,  
 RA Choy H.E.;  
 RT "Complete genome sequence of Vibrio vulnificus CMCP6";  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE016809; AA007392.1; -;  
 KW Complete proteome.  
 SQ SEQUENCE 377 AA; 42931 MW; A04DA3EA65C4679E CRC64;

Query Match 12.4%; Score 291; DB 16; Length 377;  
 Best Local Similarity 21.8%; Pred. No. 3.5e-12;  
 Matches 91; Conservative 84; Mismatches 161; Indels 82; Gaps 14;  
 QY 1 MPLSKHQIEQLSKPLSDDSICGVYKLEKSAFRPLRNEFNVAQTALRKLSONPSADERDA 60  
 16 LPISAEQPAQVD-PRSDITPQSLYRLKQRMQ-----ARNIERNALIEEPI 62  
 DB 61 LQEAELNKKWKLSDSLVEQFSKTRDELISWFAAQFLDITLESANSLWADUSEK 120  
 63 LNYA--NLWQVFLDEVPQALTSQTKLEYAWLIEALTRLD-GFKGMVAVGELATYYLEN 119  
 QY 121 HWDHLNPLVPETLTKSDDDKGEREQADAKVKAFFQVGDSESSILYAPVLQPLVGEV 180  
 120 YWADLYPM-----PDDEGLE-----TRISPIGLNGIDNEGTLIF-PLSCIPWTEGL 165

Query Match	12.1%	Score 284.5;	DB 16;	Length 373;
Best Local Similarity	24.3%;	Pred. No. 9.8e-12;		
Matches 105;	Conservative 81;	Mismatches 129;	Indels 117;	Gaps 19;

  

QY	9	EOLSKPLSDDSICGVYKLE---KSAPRLPNEFNVAQTALRKLSONPSADSRDA-LOEA	64
DB	10	EQVLAPISEDNPTGIDFREDVSPQSAFYRKQDQRMVARNA-----ERNAIISSE	58
QY	65	CL----NWKILSDLSYEQFSKTRDIELISWVAAQFLDITLLESAANSLEWLADLSEK	120
DB	59	SIHTHNMLRWVFEVEPEVLATQKQPEFVAVLIEALTRL-YGRCMGVGYKIASAFVES	117
QY	121	HWDHNLNVPVETLUKSDDDKGBKEQADAKVAFQVLGDSESSILYAPVLQLPLVG---	178
DB	118	YWDELHPM-----PDEDGIE-----TRISAITGLNGIDSEGTILIF-PLASIEPTDVG	163
QY	179	--EVTPEDFOSA-----ERKG-----EISOLKSMLTIT-----VAQERFA	211
DB	164	VEQAYAYWEYQQAELBELDDEKKRARDQGALELSRTDITVKSTDSDEFYQSLIADLEFA	223
QY	212	-----IQFMENAKRCVTQLDRLSALVSTKCHSLGSGSQSTNCFAKSLTRRVENALVHLS	265
DB	224	TGAFNFSQRLLDVGDPVP-----SSYISKRIES-----IHAALKELL	262
QY	266	GIKLAPYAAEAKTVQEQVAEASSVSEGEGLPSHMDTKHTERIPMASEQAQTVSQHLHAGNULSE	325
DB	263	GNRFKDAEV-----SIS-----DENGDEELEEAASQRAVT-----HNSSA	297

```

Qy 326 LGNNNNRDLAFHLRLREVSDFYFQSEPHSPISLELEKATRWGYSLPELLREWMSEQNG 388
Db 298 LLVTNNQSRQEAIOQ/OHVADFFRSEPHSPVSIEQIIRWCGMPLPELLAEILSD--G 355
Qy 386 DALSTIFNNAAGL 397
Db 356 DAKRSYFRLVGI 367

RESULT 3
Q93ED3 PRELIMINARY; PRT; 358 AA.
ID Q93ED3
AC Q93ED3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ImpA.
GN ImpA.
OS Rhizobium leguminosarum (biovar trifolii).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
NCBI_TaxID=386;
RN [1]
RP SEQUENCE FROM N.A.
RA Bladigroen M.R., Badelt K., Stronk O.P., Lugtenberg E.J.J.,
RA Spaink H.P.;
RT "An avirulence locus of a symbiotic Rhizobium leguminosarum strain is
RT involved in temperature-dependent protein secretion.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF361470; AAL17808.1; -.
SQ SEQUENCE 358 AA; 39039 MW; 9F5C107C58D22395 CRC64;

Query Match 10.8%; Score 252; DB 2; Length 358;
Best Local Similarity 23.9%; Pred. No. 1.7e-09;
Matches 96; Conservative 71; Mismatches 170; Indels 64; Gaps 14

Qy 8 IEQLSKPLSDPSICGVYKLKLEKSAFRLRNEFNVQATLRKLSQNPASDERDALQEAQIN 67
Db 3 VREIIDPLEQNHPCGENVR-SNTAFREIYYRIKDARNAARTAEERSIIPGTTIIAPA--- 58
Qy 68 KKKILSDLSYEQFSKTRTDELISNFWAAQFLDITLESANSLLEWAD-----LSEK 120
Db 59 -WHDVSNGLQLSKSKDLEVLAWLAERQR-----RGFSLHVVYVATVSLIDK 109
Qy 121 WHDLNLPVLPVETLKSDDDKEREQADAKYKFAFFOLVGDSESSILYAPVILQLPLVG-- 178
Db 110 HFDALHST-----GGGD-----VEERFAPFAGLNGVGEGTLLQAIRLTSIIPGKK 155
Qy 179 --EVTFDFQSAERKGISOLKSMITTVQAQERFAIQKMNAKRVTQDLRLSLAVSTK 236
Db 156 FAPQSLWDFQUSQRPNTERKLEQQAAAGVAQMSAYLDVLTGCIAAFDRMVEILDEH 215
Qy 237 CHSLGSQSTNFGFAKSLITRVENALVHLSGIKLAPKAEAKTVEQVEAVSSVSEGELPSHM 296
Db 216 CGDQAPPSSN--TRNVLQEAASAIRMLAGIE-AP-TEAPLAAPQALASQSGSE-PLPA-S 268
Qy 297 DTKHTERTPMASEQAQTVSQHLHAGNLSLSEGLNNNNRDLAFHLLREYSVDYFROQSEHSP 356
Db 269 EFDGSEALGIPVSAEMI-----RSREAPELLIAVARYFRRTPEHSP 311
Qy 357 ISFLEKAIKRWGYSLSPELLREMMXSENGDALSTIFNNAAGL 397
Db 312 ISMTETIVRGRMDFFPELLAEALLPEQ--QTRNAVLTAAGI 350

RESULT 4
Q8UVT6 PRELIMINARY; PRT; 351 AA.
ID Q8UVT6
AC Q8UVT6;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein Atu4343.

```



ATU4343 OR AGR\_L\_1042.  
 Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.  
 NCBI\_TaxID=176299;  
 [1]  
 SEQUENCE FROM N.A.  
 MEDLINE=21608550; PubMed=11743193;  
 Wood D.W., Setubal J.C., Kaul R., Morks D.E., Kitajima J.P.,  
 Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,  
 Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,  
 Chapman P., Clendinning J., Deatherage G., Gillet W., Grant C.,  
 Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,  
 Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,  
 Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,  
 Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
 Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
 Nester E.W.;  
 "The genome of the natural genetic engineer Agrobacterium tumefaciens  
 C58.";  
 Science 294:2317-2323(2001).  
 [2]  
 SEQUENCE FROM N.A.  
 MEDLINE=21608551; PubMed=11743194;  
 Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,  
 Quorlto B., Goldman B.S., Cao Y., Askenzaki M., Halling C., Mullin L.,  
 Houliet K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,  
 Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,  
 Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,  
 Cielo C., Slater S.;  
 "genome sequence of the plant pathogen and biotechnology agent  
 RT Agrobacterium tumefaciens C58.";  
 Science 294:2323-2328(2001).  
 DR EMBL: AE009363; AAL45137.1; -  
 DR EMBL: AE008251; AAK89037.1; -  
 DR PIR: AC3090; AC3090.  
 DR PIR: G98196; G98196.  
 KW Hypothetical protein; Complete proteome.  
 SQ HYPOTHEICAL 351 AA: 38490 MW: 85460P21E10C61P07 CFC64:

```

Query Match      10.3%; Score 241.5; DB 16; Length 351;
Best Local Similarity 23.6%; Pred. No. 9e-09;
Matches 96; Conservative 64; Mismatches 162; Indels 85; Gaps 15;

QY      5 KHOIEQLSKPLSDDSIGGVVLKLE--KSAFRPLRNEFNVAQTALRKLQNPSADERDAL 61
Db      7 KRRIEFLGN-----CGDIRNDSRTREIYRKDERNQAR-----AEEAAS 48

QY     62 QEACL---NRKKWILSDSLYFQSKTTRDIELISWFAAQFLDITLESANSLWEWLADLS 118
Db     49 PQDNLKISSWSDSVNGLQIITYSESXQVEILAWLAELSLR-LRGPHGLREIYELCGDLF 107

QY    119 EKHWDLNLPVLPETLKSDDDKGKERQAAKVAFQOLVGDSESSILYAPVLQLPV- 177
Db    108 YNHWSLRS:-:-----SDND-----BEKFAPFAGLNGIGSEGT-LVQPLRLASLIP 152

QY    178 ----GEVTFDFOSAERKGHISQLKSLMTTVAQERPAIQKMNAKRCVTLQDRLSALV 233
Db    153 GKGFEHSLWDFQLAQRPNESKREELRYRTASEAGVAAMSHSLAAVNTCLSSFDAITAVL 212

QY    234 STKCHSLGSGSTWFGFAKSLLTRVENALVHLSGIKLAPKBAKTV---EOEVAESSVSEG 290
Db    213 SERCGQAAPPSSN-----IRNTLIEAA-----AAIRTLGGRDOEPAP----- 249

QY    291 ELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLSELGNLNMNRDLAPHLLREYSDVPRQ 350
Db    250 -----VEQTPTA--AGTDESQCSAARTSPASPEGISSDEAFETLLSVARFRR 297

QY    351 SEPHSPISFLEKAIKRNQVYLSLBELLREMMSEQNGDALSTIFNAGL 397

Db    298 TEPHSPISLSEITLVRRGRMDFSELAEELLPET--QARNAVLTAAGI 342

```

[illegible]

```

S2                PRELIMINARY;      PRT;   337 AA.
S2:
AR-2002 (TrEMBLrel_20, Created)
AR-2002 (TrEMBLrel_20, Last sequence update)
UN-2003 (TrEMBurel_24, Last annotation update)
theoretical protein Rsp0759.
759 OR RS01949.
tonia solanacearum (Pseudomonas solanacearum).
mid megaplasmid.
emia; Proteobacteria; Betaproteobacteria; Burkholderiales;
nolderiaceae; Ralstonia.
TaxID=305;
ENCE FROM N.A.
IN=GMI1000;
INE=21681879; PubMed=11823652;
onabut M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
t M., Billault A., Brottier P., Camus J.-C., Cattolico L.,
indier C., Choisme N.; Claudiel-Renard C., Cunac S., Dewange N.,
in C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
ier P., Thibault P., Whalen M., Wincker P., Levy M.,
senbach J., Boucher C.A.;
one sequence of the plant pathogen Ralstonia solanacearum.";
re 435+497-502(2002),
;; AL44680; CAD17910.1; -.
GO:0046821; C:extrachromosomal DNA; IEA.
imid; Hypothetical protein; Complete proteome.
ENCE    337 AA; 37094 MW; C1EE68258BB39397 CRC64;
atch          9.9%; Score 232.5; DB 16; Length 337;
cal Similarity 24.9%; Pred No.3.6e-08;
99; Conservative 60; Mismatches 127; Indels 111; Gaps 18;
38 EFNVQATLRKLSQNPSADERDALQFACLNKWIKSLDSLYEQFSKTRDILISWFAAQ 97
|||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||
6 EFTIQEARR--FDPSLDQGDWVTEIKEADWRVVSESTALLQKRDXDLRAAWLTEA- 62
98 FLDDTTLESAANSLEWLADISEKHHDHNLVPVLVTETLKSDDDDKGEREQADAKVKAFQL 157
||||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||
63 LSKEGFGAGREGYELIAGLCQYWEHLFP-LP-----EPDPPEARITGSMAWL 109
158 VGDSESSIIYAPVLPLPV---GEVTFDFPQSA-----ERKGESIQSKSMLT 202
|||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||
110 AGRSSQL-----IREVPLVEAGNGFNVDWEVATNLEAIRDPDQADELSRGK---- 159
203 TTVAQRFRPAIQFWENAKR-----CVTLDRLSALVSTKCHSLGSGOSTN 246
|||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||
160 --VTPOQF-----EAARRASSAFYTYLNDWGCGAALLRLSVLDGRA---GDHAPS 208
247 FGPAKSLLTRVENALVHLSIGKL---APKAETKVQEVAESSVSSELPSHWDTKHI 301
|||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||
209 FRAREALQAV-RALVERGGGTWKPSAVEKATAQAQTAAQ----- 250
302 ERIPMASEQAQTVSQHLHAGNLSEIGNLNMMNRDLAFHLLREVSDYFROSEPHSPIBFL 361
|||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||
251 ---PVAEAVPDIV-----VGPIR--TRAQALNQLREVAEFFROTPEHPSPVAYLA 294
362 EKAIIRGYLSPELLREVMXSEQNGDA-LSTIFNAAGL 397
|||:::||:::||:::||:::||:::||:::||:::||:::||
295 ARAAKMGNDPLHANVRVVKO---DATLSQIEELLGL 328
PRELIMINARY;      PRT;   366 AA.

```

RESULT 6		
Q911B9		
ID Q911B9	PRELIMINARY;	PRT; 366 AA.
AC Q911B9;		
DT 01-MAR-2001	(TrEMBLrel. 15,	Created)
DT 01-MAR-2001	(TrEMBLrel. 15,	Last sequence update)
DT 01-JUN-2003	(TrEMBLrel. 14,	Last annotation update)
DE DE Hypothetical protein P42360.		
GN P42360.		
DN GN2360.		





```
O86A36 PRELIMINARY; PRT; 2245 AA.
ID Q86A36
AC Q86A36
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Similar to Dictyostelium discoideum (Slime mold). myosin IJ heavy chain.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RX MEDLINE=22092622; PubMed=12097910;
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P., Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K., Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.; "Sequence and analysis of chromosome 2 of Dictyostelium discoideum."; Nature 418:79-85(2002).
RL [2]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Baumgart C.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC115607; AA052027.1; -.
DR GO; GO:0016459; C:myosin; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR InterPro; IPR002710; DIL.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR004009; myosin_N.
DR Pfam; PF01843; DIL; 1.
DR Pfam; PF00612; IQ; 6.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; myosin_N; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD003376; DIL; 1.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 3.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS00096; IQ; 3.
DR SEQUENCE 2245 AA; 258682 MW; 4DA0956515A7685 CRC64;

Query Match 6.4%; Score 151; DB 5; Length 2245;
Best Local Similarity 23.6%; Pred. No. 0.26;
Matches 81; Conservative 57; Mismatches 143; Indels 62; Gaps 13;

QY 23 VYLKLESAFPLRNEFNVAOTLRKLSNPSPADRDALQEAELKWKILSDSLYEQFSK 82
DB 964 IQLRAEARSLTQVQKQKLEKLEELQWLRTSAKRKQ---LEDQKVASDTWISLS 1020
QY 83 TTRDIEL-ISWFAAQLDITLESANSL-EWLADLSEK-----HWDHNPVLPVETLK 135
DB 1021 NNDHLEQLSBIQLKQSLDKNSSQQLSECLSKLEEQQLDHDSSKLKLEKDLSD 1080
QY 136 SDDDKGKREQADAKVAFQVLGDSSESSILYAPVLQPLVGVTFP----FDFQSAER- 190
DB 1081 QHDSIEKLQSFNETEQQLQFQKQSEBLSKLSKTQ-----QLDFNKGEFDRLSQERD 1135
QY 191 -----KGEISOLKSMLTITVAQERFAIQFVENAKRCVTLDRLSALVSTKCHSLGQS 244
DB 1136 TDNTNNQLEIQLK-K-ANSTLEEDYFSLGIRDNLERQVLELRDENQILERLDSLGQQS 1194
QY 245 TNF--GFA-----KSLITVENALVHLSGKLAPKAEK-----TVE 279
DB 1195 SQFQGAALQKQLEQVQSEQLKLSSEKLGSEEAQKQINOLELTDHKSQKIQ 1254
QY 260 QVAVSSVS-----EGELPFSMDTK-----HIERIPMAEQAQ 312
DB 1255 LCLTQSQNEKIKKLGKLEEQYQDEKQQLQBELRIKQSKQSV 1297
```

```
RESULT 11
Q86IW2 PRELIMINARY; PRT; 959 AA.
ID Q86IW2
AC Q86IW2
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Hypothetical protein.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RX MEDLINE=22092622; PubMed=12097910;
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P., Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K., Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.; "Sequence and analysis of chromosome 2 of Dictyostelium discoideum."; Nature 418:79-85(2002).
RL [2]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Baumgart C.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC116986; AA051892.1; -.
DR Hypothetical protein.
DR SEQUENCE 959 AA; 110909 MW; 151551A9D7D72335 CRC64;

Query Match 6.1%; Score 143.5; DB 5; Length 959;
Best Local Similarity 19.7%; Pred. No. 0.26;
Matches 88; Conservative 78; Mismatches 173; Indels 107; Gaps 16;

QY 41 VAQTLRKL-SQNSPADE--RDA--LQEAELKWKILSD-----SLYEQFS 81
DB 498 VSDSASKLDRQADKAIQATDLQALDKAMKDIQDLNKHHSIEDRYNKKQLEERS 557
QY 82 KTRDIELISWFAAQLDITLESANSL-EWLADLSEKHWDLNPVLPVETLKSDDDKG 141
DB 558 RIKSEFQLIANNEKEFL-----DKIEQLETFIQEHRSTINRNEKQTKW-EELN 607
QY 142 KE-----REQADAKVAFQVLGDSSESSILYAPVLQPLVGEVTFDFQSAERKGEISQ 196
DB 608 KEIAHLQQRCDAEIR-----NDQLSSI---PQATEPLIKQIESIQEQFNERQSTWET 658
QY 197 LKSMLTITVAQERFAIQFVENAKRCVTLDRLSALVSTKCHSLGQSQTNEGFAKSLITR 256
DB 659 LEKQNNQJREELKAEHVVRNQBELSELEL-----TLKQLLES-----KIDKK 708
QY 257 VENALVHLSGKLAPKAEAKTVEQVAVSSVSEGLPSHMDTKHI--BRIPMAS----- 308
DB 709 LKSTITELQQLKSIQSDHLSKIQDLESSSSTLQSKIDQHLANTISILEERLTLSNKKDL 768
QY 309 -----EQQTVSQHLHAGNLSELGNLNNNNRDLAPHL 340
DB 769 EDKQKEKELYKETSNFLKQQLQLLQQLQQLQQLQQLQQLQQLQQLQQLQQLQQLQ 828
QY 341 LREVSDFRQSEPHSP-----ISFLEKAIKRWGYS-----LP---ELIREMMGE 382
DB 829 LNSDTDKINNGISITFPQQQNKGLFSGLKTPSNHDFLNHVSNGLLPASLEYLQSSLSQ 888
QY 383 QNGDALSTTFNAGNLHLDQVLLPEV 408
DB 889 KEGEAMSLQAVQVQSLNHSKLEDEL 914

RESULT 12
O02239 PRELIMINARY; PRT; 884 AA.
ID O02239
AC O02239; O02269;
DT 01-JUL-1997 (Tremblrel. 04, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
```



```

Db 294 EYVVRHPSPPVLLKRAKTLVADPAEIVRNLPD-----GISQFETLRG 340
QY 406 PE 407
Db 341 PE 342

RESULT 14
Q8MPV7 PRELIMINARY; PRT; 1534 AA.
AC Q8MPV7;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Hypothetical protein T23F2.2a.
GN T23F2.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Feloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
EX MEDLINE=99069613; PubMed=9851916;
RA Waterston R.;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Du Z.;
RT "The sequence of C. elegans cosmid T23F2.";
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; U99649; AAM69069.1; -.
DR WormPep; T23F2.2a; CE31245.
KW Hypothetical protein.
SQ SEQUENCE 1534 AA; 177345 MW; B82EF064300CCE52 CRC64;

Query Match 6.0%; Score 141; DB 5; Length 1534;
Best Local Similarity 20.9%; Pred. No. 0.75; Indels 178; Gaps 27;
Matches 115; Conservative 90; Mismatches 168;

QY 9 EQLSKPLSDSDSCGYYLKLKSAFRLNPNFVAQTALRKLSONPSADERDALQACLNK 68
Db 619 QQLSEHSD-----FLSTQQT-LHDLKKEYSAKNTTL-----VDKFEVEEILLAK 663
QY 69 WKILSDSIYEQFSK-----TTTDIELISWFAVAAQFLDITLESANSLWLAADLSEKHWDH 124
Db 664 TELV-DALTQKLENIRKDOTRELS-----LKQSERDQYKKSLEEMTFIAEK----- 708
QY 125 LNPVLPPVETLKSDDDKGEREQADAKVK---AFFQ-----LVGDS----- 161
Db 709 -VPILAEAILQLSKDK---NEITARLKHQDQYFEDELAKLLNDSMNKIKERDDYLTEHI 763
QY 162 --ESSILYAPVLQPLVGEVTFDFQSAER-KGEISQLKSMLTITVAQERFAIQPKMEN 218
Db 764 RANESMI-----ERLKEISGLKKLENQKQAHQ---KSEL 798
QY 219 AKRCVTOLDRLSALVS-----TKCHSLGSQTNFGFAK----- 251
Db 799 EKLLSSIDHVSQLOSRVNHQSRDVEQCAIPQINKYVGCKPNVKNKETIIEKALFDEN 858
QY 252 -----SLTRVENALVHL-----SGIKLAPKAEAKTVSQEVAESS 286
Db 859 BERLRICKAELETTTRQVTLQKLVLSIIQQOSSQKIKGRIAVVEDSNKNTVHTEDLESK 918
QY 287 VSEGLPSPHMDTKHIERIPMAEQTAQTVSOHLHAGNLSLGN-----LNNMNR 334

```

```

Db 919 MKEVEL---KNTLMERID-SLEAEFRVASSIEKRIQKLVNEFDNLKQKLDNDSNYSK 974
QY 335 DLAF-----HLLREVSDYFRQSEPHSPISFLLEKAIKRWGYLSLPE-LLREMMGEQN-GD 386
Db 975 EKQWLQWRISNLEKNSELQIQPSS-----EKSLE-----SLNKGTLRKTMTSEPDG 1024
QY 387 ALSTIFNAAGLHLDQVLLPEVSTPTVGIESPQT-POAKPSVSDPSRVEEHVQSOTSPVDT 445
Db 1025 DMSTEGDGASTNESADFMVESVTAPVLSKSPQPSQLADVLNLYRSOLEQV-LTEIBE 1083
QY 446 QSKQDKQKPOSS 456
Db 1084 EAAKQEEFQMS 1094

RESULT 15
Q8MPV6 PRELIMINARY; PRT; 1538 AA.
AC Q8MPV6;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Hypothetical protein T23F2.2b.
GN T23F2.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Feloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
EX MEDLINE=99069613; PubMed=9851916;
RA Waterston R.;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Du Z.;
RT "The sequence of C. elegans cosmid T23F2.";
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; U99649; AAM69070.1; -.
DR WormPep; T23F2.2b; CE31246.
KW Hypothetical protein.
SQ SEQUENCE 1538 AA; 177862 MW; E6E8E4C16C941A85 CRC64;

Query Match 6.0%; Score 141; DB 5; Length 1538;
Best Local Similarity 20.9%; Pred. No. 0.76;
Matches 115; Conservative 90; Mismatches 168; Indels 178; Gaps 27;

QY 9 EQLSKPLSDSDSCGYYLKLKSAFRLNPNFVAQTALRKLSONPSADERDALQACLNK 68
Db 619 QQLSEHSD-----FLSTQQT-LHDLKKEYSAKNTTL-----VDKFEVEEILLAK 663
QY 69 WKILSDSIYEQFSK-----TTTDIELISWFAVAAQFLDITLESANSLWLAADLSEKHWDH 124
Db 664 TELV-DALTQKLENIRKDOTRELS-----LKQSERDQYKKSLEEMTFIAEK----- 708
QY 125 LNPVLPPVETLKSDDDKGEREQADAKVK---AFFQ-----LVGDS----- 161
Db 709 -VPILAEAILQLSKDK---NEITARLKHQDQYFEDELAKLLNDSMNKIKERDDYLTEHI 763
QY 162 --ESSILYAPVLQPLVGEVTFDFQSAER-KGEISQLKSMLTITVAQERFAIQPKMEN 218
Db 764 RANESMI-----ERLKEISGLKKLENQKQAHQ---KSEL 798

```

```

Qy 219 AKRCVTOLDRLSALVS-----TKCHSLCSOSTNFGFAK----- 251
Db 799 EKLLSSIDHVSQSRVNHRSQRDEVCQAIPRQINKYVGCKPNVKNKETTIEKGALEFDEN 858
Qy 252 -----SLLTRVENALVHL-----SGIKLAPKAEAKTVECEVAESS 286
Db 859 EERLRICKAELETTTRQVTVLQCKLVSIIOQSSSQIKKRIAVVEDSNKNTVHTEDLESK 918
Qy 287 VSECELP SHMDTKHIERIPWASEQAOQTVSQHLHAGNLSELGN-----LNNMNR 334
Db 919 MKEVEL---XNTELMERID-SLEAERFVASSIEKSRIQKLVNBEFDNLKQKLDNDMGNYSK 974
Qy 335 DLAF-----HLUREVSDYFRQSEPHSPISFLLEKAIKWYLSLPE-LLEMMVSEON-GD 386
Db 975 EKQWLQWRISNLEKDNSELQKIQPSS-----EKSLE-----SLNKGTLRKTMSEPDFGD 1024
Qy 387 ALSTIFNAAGLNHLQVLLPEVSTPTVGIESPOT-POAKPSVSDPRSVEEHVQSOTSPVDI 445
Db 1025 DMSTEGGASTNESADFMVESVTAPVLSLKSPOFFSQADVLNLRSDLEQV-LTEIEEP 1083
Qy 446 QSKODQKPQSS 456
Db 1084 EAAKQEPQMS 1094

```

Search completed: July 3, 2004, 02:46:15  
Job time : 76 secs

